Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/AU04/001844

International filing date: 24 December 2004 (24.12.2004)

Document type: Certified copy of priority document

Document details: Country/Office: AU
Number: 2003907150
Filing date: 24 December 2003 (24.12.2003)

Date of receipt at the International Bureau: 25 January 2005 (25.01.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

PCT/AU04/001844

Australian Government

Patent Office
Canberra

I, LEANNE MYNOTT, MANAGER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. 2003907150 for a patent by G2 THERAPIES LTD as filed on 24 December 2003.

WITNESS my hand this
Seventeenth day of January 2005

LEANNE MYNOTT
MANAGER EXAMINATION SUPPORT
AND SALES

AUSTRALIA

Patents Act 1990

G2 Therapies Ltd

PROVISIONAL SPECIFICATION

Invention Title:

Transgenic Mammals

The invention is described in the following statement:

Transgenic Mammals

FIELD OF THE INVENTION

The present invention relates to transgenic non-human mammals comprising a polynucleotide encoding a human C5aR. The invention also relates to use of the transgenic non-human mammals in methods of screening for agonists and antagonists of human C5aR and for testing efficacy of C5aR agonists and antagonists in various animal models of disease.

BACKGROUND OF THE INVENTION

Proteolysis of each of the complement proteins C3-C5 gives rise to aminoterminal cationic fragments with signalling molecules called anaphylatoxins. The most potent of these, C5a, elicits the broadest responses. Considering the components of the inflammatory response as margination and infiltration of leukocytes, release of granule-bound proteolytic enzymes, production of activated oxygen and nitrogen-derived radicals, changes in blood flow and capillary leakage, along with the ability to contract smooth muscle, the C5a molecule is the "complete" pro-inflammatory mediator. At sub-nanomolar to nanomolar levels, the C5a molecule elicits chemotaxis of all myeloid lineages (neutrophils, eosinophils and basophils, macrophages and monocytes), and causes vascular permeability which is markedly potentiated by prostaglandins and circulating leukocytes. Higher nanomolar concentrations elicit degranulation and activation of NADPH oxidase. This breadth of bioactivity contrasts with other inflammatory mediators. C5a has been implicated in the pathogenesis of rheumatoid arthritis, psoriasis, sepsis, reperfusion injury, and adult respiratory distress syndrome.

The activities of C5a are mediated by the binding of the C5a to its receptor (C5aR). C5aR belongs to the family of seven transmembrane G-protein-coupled receptors. C5aR is a high affinity receptor for C5a, with a Kd of ~ 1nM, and is located on a number of different cell types including leukocytes. The number of receptors per cell is extremely high, up to 200,000 sites per leukocyte. Biological activation of the receptor occurs over the range that saturates binding.

C5aR comprises an extended N-terminal extracellular domain. This large N-terminal domain is typical of G-protein coupled receptors which bind peptides including the IL-

8 and fMet-Leu-Phe (FMLP) receptor families. The C5aR structure conforms to the seven transmembrane receptor family, with the extracellular N-terminus being followed by seven transmembrane helices connected by interhelical domains alternating as intracellular and extracellular loops, and ending with an intracellular C-terminal domain.

Agonists of C5aR are useful for therapeutic purposes, for example, in defence against bacterial infection, to stimulate immunoregulatory effects of C5a, and to treat cancers, immunodeficiency diseases and severe infections.

Antagonists of C5aR are also useful therapeutic agents for example, for treating inflammatory diseases and autoimmune disorders. For example, antagonists of C5aR are useful in the treatment of asthma, bronchial allergy, chronic inflammation, systemic lupus erythematosis, vasculitis, rheumatoid arthritis, osteoarthritis, gout, some auto-allergic diseases, transplant rejection, inflammatory bowel disease (for example, ulcerative colitis), in certain shock states, myocardial infarction, and post-viral encephalopathies. To this end, C5aR peptide antagonists and anti-C5a receptor antibodies have been previously described. For example, WO95/00164 describes antibodies directed against an N-terminal peptide (residues 9-29) of the C5a receptor.

Currently, alternative and/or improved C5aR antagonists and antagonists are desirable, as are improved methods of screening for C5aR antagonists and antagonists.

In vitro screening methods for detecting of C5aR agonists/antagonists are known in the art. For example, chemotaxis assays can be used to assess the ability of an antibody or functional fragment thereof to block binding of a ligand to C5aR and/or inhibit function associated with binding of the ligand to the receptor. These assays are based on the functional migration of cells *in vitro* induced by a compound. Chemotaxis can be assessed by any suitable means, for example, in an assay utilizing a 96-well chemotaxis plate, or using other art-recognized methods for assessing chemotaxis. For example, the use of an *in vitro* transendothelial chemotaxis assay is described by Springer *et al.* (Springer *et al.*, WO 94/20142, published Sep. 15, 1994; see also Berman et al., Immunol. Invest. 17: 625-677 (1988)). Migration across endothelium into collagen gels has also been described (Kavanaugh *et al.*, J. Immunol., 146: 4149-4156 (1991)).

In vitro assays suffer from the limitation that they do not provide any information concerning the pharmacokinetics or bioavailability of the compound being tested or the true *in vivo* consequence of receptor inhibition. For these reasons, ligands identified by screening on cells *in vitro* must in any case be subjected to tests on animals, particularly animal models of disease to test the efficacy of an antagonist or agonist.

Improved screening methods that allow identification of C5aR agonists/antagonists in an *in vivo* environment are therefore desirable.

SUMMARY OF THE INVENTION

The present inventors have found that a number of C5aR antagonists react with human C5aR but not C5aR from other species. For example, monoclonal antibodies MAb 7F3, MAb 6C12 and MAb 12D4 (described in PCT/AU03/00084) bind to human C5aR but do not bind to mouse or baboon C5aR. This brings to light the need for an *in vivo* screening and validation system that is capable of detecting and/or validating agonists/antagonists that are specific for human C5aR.

In work leading up to the invention, the present inventors found that chemotaxis of murine cells engineered to express human C5aR is inhibited by anti-human C5aR antibodies. This finding, coupled with the knowledge that murine C5a binds to human C5aR with high affinity, indicates that human C5aR is compatible with C5aR signalling machinery in other mammalian systems. The present inventors have therefore taken steps to develop a transgenic non-human mammal that expresses a human C5aR and is useful for screening for agonists/antagonists of C5aR, particularly those agonists/antagonists specific for human C5aR.

Accordingly, in a first aspect the present invention provides a transgenic non-human mammal comprising a polynucleotide encoding a human C5aR or a fragment thereof.

In a preferred embodiment of the first aspect, the polynucleotide encodes a fragment of human C5aR. Preferably, the fragment encompasses at least one domain of C5aR.

The various domains of human C5aR suitable for use in generating a transgenic mammal of the invention are listed in Table 1.

Table 1:

amino acids	1 - 37	extracellular domain - N-terminus
amino acids	38 - 60	transmembrane domain
amino acids	61 - 71	intracellular domain
amino acids	72 - 94	transmembrane domain
amino acids	95 - 110	extracellular domain - extracellular loop 1
amino acids	111 - 132	transmembrane domain
amino acids	133 - 153	intracellular domain
amino acids	154 - 174	transmembrane domain
amino acids	175 - 200	extracellular domain - extracellular loop 2
amino acids	201 - 226	transmembrane domain
amino acids	227 - 242	intracellular domain
amino acids	243 - 265	transmembrane domain
amino acids	266 - 282	extracellular domain - extracellular loop 3
amino acids	283 - 303	transmembrane domain
amino acids	304 - 350	intracellular domain - C-terminus.

In a preferred embodiment of the invention, the fragment encompasses at least one of the domains listed in Table 1. In one particular embodiment, the fragment encompasses at least one extracellular domain of human C5aR. In another embodiment, the fragment encompasses two or more of the domains listed in Table 1.

In a further embodiment, the polynucleotide encodes a chimeric C5aR. Preferably, the chimeric C5aR comprises a C5aR sequence endogenous to the transgenic mammal wherein at least one endogenous domain is replaced with the corresponding domain of human C5aR. For example, the chimeric C5aR may comprise an endogenous C5aR sequence wherein at least one extracellular or intracellular domain is replaced with the corresponding human C5aR domain. In another example, the chimeric C5aR comprises intracellular domains of the endogenous C5aR and extracellular domains of human C5aR. Such a chimeric C5aR is referred to herein as "humanized C5aR".

In another preferred embodiment, the polynucleotide encodes a complete human C5aR. Preferably, the polynucleotide encodes a polypeptide comprising a sequence as shown in SEQ ID NO:1, or a biologically active variant, derivative or fragment thereof.

In a further preferred embodiment, the polynucleotide comprises a sequence as shown in SEQ ID NO:2, or a biologically active variant, derivative or fragment thereof.

The polynucleotide encoding humanized C5aR or human C5aR or a fragment thereof may be introduced into the cells of the mammal by way of any one of a number of suitable transgenic techniques known in the art. Preferably, following introduction into the cells the polynucleotide encoding human C5aR or a fragment thereof is integrated within a host chromosome.

It will be appreciated that a transgenic mammal comprising a polynucleotide encoding humanized C5aR may be generated by introducing a polynucleotide encoding a humanized C5aR into the genome of the transgenic mammal or by targeted integration of one or more human C5aR domains in to the endogenous sequence such that the human C5aR domains replace the corresponding endogenous domains in the genome of the transgenic mammal.

In a further preferred embodiment the transgenic mammal has somatic and germline cells which contain, in a stably integrated form, a polynucleotide encoding a human or humanized C5aR. In other words, it is preferred that the transgenic mammal is a "knock in" for human or humanized C5aR. In a further preferred embodiment, the transgenic mammal is homozygous for human or humanized C5aR.

In a further preferred embodiment, expression of endogenous C5aR in the transgenic animal is undetectable or insignificant. Reduction in expression of endogenous C5aR may be achieved by any suitable means. For example, the cells of the transgenic mammal may be modified so as to express an antisense nucleic acid complementary to nucleic acids encoding endogenous C5aR. Alternatively, the endogenous C5aR gene may be disrupted by homologous recombination. Preferably, the transgenic mammal is a homozygous "knock out" for endogenous C5aR.

In a preferred embodiment of the invention, the "knock out" of endogenous C5aR occurs simultaneously with the introduction of human or humanized C5aR. This is preferably achieved by replacing the endogenous C5aR coding sequence or a fragment thereof with a corresponding human C5aR coding sequence or fragment thereof by way of targeted homologous recombination. In one particular embodiment of the invention,

one or more of the domains of the endogenous C5aR is replaced with the corresponding human domain(s).

The transgenic animals of the present invention may comprise other genetic alterations in addition to the presence of a human C5aR encoding sequence. For example, the genome of the transgenic animal may be altered to affect the function of endogenous genes , contain marker genes, or other genetic alterations consistent with the methods of the present invention.

In a further preferred embodiment of the invention, the transgenic non-human mammal is selected from the group consisting of a cow, pig, goat, sheep, camel, horse, cat, dog, monkey, baboon, rabbit, guinea pig, rat, hamster and mouse. Rodents such as rats, mice and hamsters are preferred mammals. Preferably, the transgenic mammal is a mouse.

Also encompassed by the present invention are methods of identifying ligands or other substances which bind C5aR, including agonists and/or antagonists of human C5aR function, which involve the use of a transgenic mammal (or cells derived therefrom) of the present invention. For example, a putative compound is administered to the transgenic animal and a response of the transgenic animal to the putative compound is measured and compared to the response of a "normal" or wild-type mouse or, alternatively, compared to a transgenic animal control (without administration of the compound). The invention further provides agents identified according to such methods. The present invention also provides methods of identifying compounds useful as therapeutic agents for treating conditions associated with C5aR signalling.

Accordingly, in a further aspect the present invention provides a method of identifying a compound that modulates C5aR activity, the method comprising (i) administering a candidate compound to a transgenic mammal of the present invention or cells derived therefrom under conditions in which at least one phenotype associated with C5aR signalling is expressed; and (ii) monitoring development of the at least one phenotype following administration of the compound.

In a preferred embodiment the method further comprises (iii) comparing the extent of the phenotype in the transgenic mammal or cells derived therefrom to which the compound was administered relative to a control mammal or cells derived therefrom,

wherein any difference in the nature or extent of the phenotype indicates the potential of the compound to modulate C5aR activity.

In one embodiment the present invention provides a method of identifying a compound that inhibits or reduces C5aR activity, the method comprising (i) administering a candidate compound to a transgenic mammal of the present invention under conditions in which at least one phenotype associated with C5aR signalling is expressed; (ii) monitoring development of the at least one phenotype following administration of the compound; (iii) and comparing the extent of the phenotype in the transgenic mammal to which the compound was administered relative to a control mammal, wherein any inhibition or reduction in the nature or extent of the phenotype indicates the potential of the compound to inhibit or reduce C5aR activity.

In another embodiment the present invention provides a method of identifying a compound that promotes or enhances C5aR activity, the method comprising (i) administering a candidate compound to a transgenic mammal of the present invention under conditions in which at least one phenotype associated with C5aR signalling is expressed; (ii) monitoring development of the phenotype following administration of the compound; (iii) and comparing the extent of the phenotype in the transgenic mammal to which the compound was administered relative to a control mammal, wherein any enhancement in the nature or extent of the phenotype indicates the potential of the compound to promote or enhance C5aR activity.

The "control" animal employed in this context can be any other mammal that expresses the same phenotypic indictors as those expressed in the mammal on which the compound was tested (i.e. the "test" mammal). Preferably, the control and test animals express similar levels of functional human C5aR. More preferably, the control and test animals are isogenic.

The phrase "phenotype associated with C5aR signalling" is intended to encompass any visible characteristic and/or behaviour (including a clinical symptom of a disease) that is associated with a biochemical process involving C5aR signalling. The phenotype may be associated with normal or aberrant C5aR signalling. In one embodiment the phenotype is a condition that is aggravated by C5aR signalling such as an immune complex disorder, an inflammatory or allergic disease, graft rejection or cancer. In

another embodiment, the phenotype is a condition that is alleviated or abated by increased C5aR signalling such as a condition associated with immunosuppression.

It will be appreciated by those skilled in the art that any suitable phenotype associated with C5aR signalling may be monitored in the screening method of the present invention.

In one embodiment, the phenotype is leukocyte infiltration.

In another embodiment, the phenotype is asthma.

In another embodiment, the phenotype is sepsis, stroke or respiratory distress syndrome.

In another embodiment, the phenotype is inflammation or inflammatory tissue damage such as ischaemia-reperfusion injury.

In another embodiment, the phenotype is a condition selected from the group consisting of inflammatory or allergic diseases and conditions, including respiratory allergic diseases such as asthma, allergic rhinitis, hypersensitivity lung diseases, hypersensitivity pneumonitis, interstitial lung diseases (ILD) (e.g., idiopathic pulmonary fibrosis, or ILD associated with rheumatoid arthritis, systemic lupus erythematosus, ankylosing spondylitis, systemic sclerosis, Sjogren's syndrome, polymyositis or dermatomyositis); anaphylaxis or hypersensitivity responses, drug allergies (e.g., to penicillin, cephalosporins), insect sting allergies; inflammatory bowel diseases, such as Crohn's disease and ulcerative colitis; spondyloarthropathies; scleroderma; psoriasis and inflammatory dermatoses such as dermatitis, eczema, atopic dermatitis, allergic contact dermatitis, urticaria; vasculitis (e.g., necrotizing, cutaneous, and hypersensitivity vasculitis); autoimmune diseases, such as arthritis (e.g., rheumatoid arthritis, psoriatic arthritis), multiple sclerosis, systemic lupus erythematosus, myasthenia gravis, juvenile onset diabetes, nephritides such as glomerulonephritis, autoimmune thyroiditis, Behcet's disease; graft rejection (e.g., in transplantation), including allograft rejection or graft-versus-host disease; atherosclerosis; cancers with leukocyte infiltration of the skin or organs; reperfusion injury, stroke, adult respiratory distress syndrome, certain hematologic malignancies, cytokine-induced toxicity (e.g., septic shock, endotoxic shock), polymyositis,

dermatomyositis, pemphigoid, Alzheimers Disease , granulomatous diseases including sarcoidosis, immunodeficiency syndromes such as AIDS, radiation therapy, chemotherapy, therapy for autoimmune disease or other drug therapy (e.g., corticosteroid therapy), which causes immunosuppression; and immunosuppression due congenital deficiency or infectious diseases such as Severe Acute Respioratory Syndrome (SARS).

The range of candidate compounds contemplated herein include C5aR inhibitory compounds or antagonists of a biological function of C5aR. In one embodiment, the compound is selected from the group consisting of: a peptide, including a peptide derived from C5aR and capable of inhibiting, reducing or repressing a C5aR function, a C5aR dominant-negative mutant; a non peptide inhibitor of C5aR; an antibody or antibody fragment which binds to C5aR and inhibits a C5aR function; a small organic molecule, and nucleic acid, including nucleic acid encoding said peptide derived from C5aR or said non-C5aR peptide inhibitor, an antisense nucleic acid directed against C5aR-encoding mRNA, or an anti-C5aR ribozyme, or a small interfering RNA (RNAi) that targets C5aR gene expression.

In one embodiment, the subject method further comprises formulating the identified compound for administration to a non-human animal or a human. The formulations can be suitable for administration by injection by a subcutaneous, intravenous, intranasal, or intraperitoneal route. Alternatively, they can be suitable for oral administration in the form of feed additives, tablets, troches, etc.

In another aspect, the present invention provides a compound identified by a method of the present invention.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Diagram showing design of targeting construct for generation of human C5aR knock-in mice.

Figure 2: Diagram showing transgenic mouse C5aR locus after deletion of the PGKneo gene by Cre recombinase.

Figure 3: Sequence of mouse mouse C5aR(C5r1) gene locus and flanking DNA.

Figure 4: Sequence of human C5aR cDNA.

Figure 5: Amino acid sequence of human C5aR.

Figure 6: Diagram showing selected restriction sites and mouse C5aR (C5r1) gene exons.

DETAILED DESCRIPTION OF THE INVENTION

General techniques and Definitions

Unless otherwise indicated, the recombinant DNA techniques utilized in the present invention are standard procedures, well known to those skilled in the art. Such techniques are described and explained throughout the literature in sources such as, J. Perbal, A Practical Guide to Molecular Cloning, John Wiley and Sons (1984), J. Sambrook et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory Press (1989), T.A. Brown (editor), Essential Molecular Biology: A Practical Approach, Volumes 1 and 2, IRL Press (1991), D.M. Glover and B.D. Hames (editors), DNA Cloning: A Practical Approach, Volumes 1-4, IRL Press (1995 and 1996), and F.M. Ausubel et al (Editors), Current Protocols in Molecular Biology, Greene Pub. Associates and Wiley-Interscience (1988, including all updates until present) and are incorporated herein by reference. In particular, these documents describe in detail methods of transcribing or replicating nucleic acid molecules and suitable conditions required therefor.

Definitions

As used herein, a "ligand" of a C5aR protein refers to a particular class of substances which bind to a mammalian C5aR protein, including natural ligands and synthetic and/or recombinant forms of natural ligands. In a preferred embodiment, ligand binding of a C5aR protein occurs with high affinity.

As used herein, an "antagonist" is a substance which inhibits (decreases or prevents) at least one function characteristic of a C5aR protein such as a binding activity (e.g.,

ligand binding, promoter binding, antibody binding), a signaling activity (e.g., activation of a mammalian G protein, induction of rapid and transient increase in the concentration of cytosolic free calcium) and/or cellular response function (e.g., stimulation of chemotaxis, exocytosis or inflammatory mediator release by leukocytes). The term antagonist encompasses substances which bind receptor (e.g., an antibody, a mutant of a natural ligand, small molecular weight organic molecules, other competitive inhibitors of ligand binding), and substances which inhibit receptor function without binding thereto (e.g., an anti-idiotypic antibody).

As used herein, an "agonist" is a substance which promotes (induces, causes, enhances or increases) at least one function characteristic of a C5aR protein such as a binding activity (e.g., ligand, inhibitor and/or promoter binding), a signaling activity (e.g., activation of a mammalian G protein, induction of rapid and transient increase in the concentration of cytosolic free calcium) and/or a cellular response function (e.g., stimulation of chemotaxis, exocytosis or inflammatory mediator release by leukocytes). The term agonist encompasses substances which bind receptor (e.g., an antibody, a homolog of a natural ligand from another species), and substances which promote receptor function without binding thereto (e.g., by activating an associated protein). In a preferred embodiment, the agonist is other than a homolog of a natural ligand.

By "transgenic mammal" is meant a mammal (e.g, mouse, rat, hamster, etc.), having a non-endogenous (i.e., heterologous) nucleic acid sequence present as an extrachromosomal element in a portion of its cells or stably integrated into its germ line DNA (i.e., in the genomic sequence of most or all of its cells). Heterologous nucleic acid is introduced into the germ line of such transgenic animals by genetic manipulation of, for example, embryos or embryonic stem cells of the host animal.

The term "biologically active" or "functional", when used herein as a modifier of human C5aR, refers to a polypeptide that exhibits at least one of the functional characteristics attributed to native human C5aR, such as the ability to function as a C5a receptor or to bind to one or more natural ligands of human C5aR.

A "knock-out" of a gene means an alteration in the sequence of the gene that results in a decrease of function of the target gene, preferably such that target gene expression is undetectable or insignificant. A knock-out of an endogenous gene means that function of the gene has been substantially decreased so that expression is not detectable or only

present at insignificant levels. "Knock-out" transgenics can be transgenic animals having a heterozygous knock-out of a gene or a homozygous knock-out of a gene. "Knock-outs" also include conditional knock-outs, where alteration of the target gene can occur upon, for example, exposure of the animal to a substance that promotes target gene alteration, introduction of an enzyme that promotes recombination at the target gene site (e.g., Cre in the Cre-lox system), or other method for directing the target gene alteration postnatally.

A "knock-in" of a target gene means an alteration in a host cell genome that results in altered expression (e.g., increased (including ectopic)) of the target gene, e.g., by introduction of an additional copy of the target gene, or by operatively inserting a regulatory sequence that provides for enhanced expression of an endogenous copy of the target gene. "Knock-in" transgenics of interest for the present invention are transgenic animals having a knock-in of a human or humanized C5aR. Such transgenics can be heterozygous knock-in for the human or humaized C5aR gene or homozygous for the knock-in of the human or humanized C5aR gene. "Knock-ins" also encompass conditional knock-ins.

The term "ES cell" as used herein refers to pluripotent embryonic stem cells and to such pluripotent cells in the very early stages of embryonic development, including but not limited to cells in the blastocyst stage of development.

By "construct" is meant a recombinant nucleic acid, generally recombinant DNA, that has been generated for the purpose of the expression of a specific nucleotide sequence, or is to be used in the construction of other recombinant nucleotide sequences.

By "operably linked" is meant that a DNA sequence and a regulatory sequence are connected in such a way as to permit gene expression when the appropriate molecules (e.g., transcriptional activator proteins) are bound to the regulatory sequence.

By "operatively inserted" is meant that a nucleotide sequence of interest is positioned adjacent a nucleotide sequence that directs transcription and translation of the introduced nucleotide sequence of interest (i.e., facilitates the production of, e.g., a polypeptide encoded by a human C5aR sequence).

The term "corresponds to" is meant homologous to or substantially equivalent to or functionally equivalent to the designated sequence.

The term "transgenic gene construct" refers to a nucleic acid molecule, e.g., a vector, containing the subject polynucleotide, e.g., the human C5aR polynucleotide or fragment thereof, operably linked in a manner capable of expressing the polynucleotide in a host cell. As used herein, the term "polynucleotide" encompasses deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). As used herein the term also encompasses analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides.

By "genomic sequence" is meant a sequence having non-contiguous open reading frames, where introns interrupt the protein coding regions. It may further include the 3' and 5' untranslated regions found in the mature mRNA. It may further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, etc., including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' or 3' end of the transcribed region. The genomic DNA may be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence.

Throughout this specification the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

Transgenic Animals

Techniques for producing transgenic animals are well known in the art. A useful general textbook on this subject is Houdebine, Transgenic animals – Generation and Use (Harwood Academic, 1997) – an extensive review of the techniques used to generate transgenic animals from fish to mice and cows. Of particular interest in the context of the present invention are transgenic non-human mammals such as cows, pigs, goats, horses, etc., and particularly rodents, e.g. rats, mice, hamsters, etc. Preferably, the transgenic animal is a mouse.

Advances in technologies for embryo micromanipulation now permit introduction of heterologous DNA into, for example, fertilized mammalian ova. For instance, totipotent or pluripotent stem cells can be transformed by microinjection, calcium phosphate mediated precipitation, liposome fusion, retroviral infection or other means, the transformed cells are then introduced into the embryo, and the embryo then develops into a transgenic animal. In a preferred method, developing embryos are infected with a retrovirus containing the desired DNA, and transgenic animals produced from the infected embryo. In a further preferred method, however, the appropriate DNAs are coinjected into the pronucleus or cytoplasm of embryos, preferably at the single cell stage, and the embryos allowed to develop into mature transgenic animals. Those techniques as well known. See reviews of standard laboratory procedures for microinjection of heterologous DNAs into mammalian fertilized ova, including Hogan *et al.*, Manipulating the Mouse Embryo, (Cold Spring Harbor Press 1986); Krimpenfort *et al.*, Bio/Technology 9:844 (1991); Palmiter *et al.*, Cell, 41: 343 (1985); Kraemer *et al.*, Genetic manipulation of the Mammalian Embryo, (Cold Spring Harbor Laboratory Press 1985); Hammer *et al.*, Nature, 315: 680 (1985); Wagner *et al.*, U.S. Pat. No. 5,175,385; Krimpenfort *et al.*, U.S. Pat. No. 5,175,384, the respective contents of which are incorporated herein by reference.

Another method used to produce a transgenic animal involves microinjecting a nucleic acid into pro-nuclear stage eggs by standard methods. Injected eggs are then cultured before transfer into the oviducts of pseudopregnant recipients.

Transgenic animals may also be produced by nuclear transfer technology as described in Schnieke, A.E. *et al.*, 1997, Science, 278: 2130 and Cibelli, J.B. *et al.*, 1998, Science, 280: 1256. Using this method, fibroblasts from donor animals are stably transfected with a plasmid incorporating the coding sequences for a binding domain or binding partner of interest under the control of regulatory. Stable transfectants are then fused to enucleated oocytes, cultured and transferred into female recipients.

Analysis of animals which may contain transgenic sequences would typically be performed by either PCR or Southern blot analysis following standard methods.

By way of a specific example for the construction of transgenic mammals, such as cows, nucleotide constructs comprising a sequence encoding a binding domain fused to GFP are microinjected using, for example, the technique described in U.S. Pat. No.

4,873,191, into oocytes which are obtained from ovaries freshly removed from the mammal. The oocytes are aspirated from the follicles and allowed to settle before fertilization with thawed frozen sperm capacitated with heparin and prefractionated by Percoll gradient to isolate the motile fraction.

The fertilized oocytes are centrifuged, for example, for eight minutes at 15,000 g to visualize the pronuclei for injection and then cultured from the zygote to morula or blastocyst stage in oviduct tissue-conditioned medium. This medium is prepared by using luminal tissues scraped from oviducts and diluted in culture medium. The zygotes must be placed in the culture medium within two hours following microinjection.

Oestrous is then synchronized in the intended recipient mammals, such as cattle, by administering coprostanol. Oestrous is produced within two days and the embryos are transferred to the recipients 5-7 days after estrous. Successful transfer can be evaluated in the offspring by Southern blot.

Alternatively, the desired constructs can be introduced into embryonic stem cells (ES cells) and the cells cultured to ensure modification by the transgene. The modified cells are then injected into the blastula embryonic stage and the blastulas replaced into pseudopregnant hosts. The resulting offspring are chimeric with respect to the ES and host cells, and nonchimeric strains which exclusively comprise the ES progeny can be obtained using conventional cross-breeding. This technique is described, for example, in WO91/10741.

Transgenic animals comprise an exogenous nucleic acid sequence present as an extrachromosomal element or stably integrated in all or a portion of its cells, especially in germ cells. It is preferred that a transgenic animal comprises stable changes to the germline sequence. A stable change is generally achieved by introduction of the DNA into the genome of the cell. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like. During the initial construction of the animal, "chimeras" or "chimeric animals" are generated, in which only a subset of cells have the altered genome. Chimeras are primarily used for breeding purposes in order to generate the desired transgenic animal. Animals having a heterozygous alteration are generated by breeding of chimeras. Male and female heterozygotes are typically bred to generate homozygous animals.

In a preferred embodiment, the transgenic non-human mammals of the invention are produced by introducing a human C5aR transgene into the germline of the non-human animal. Embryonal stem cell (ES) are the primary type of target cell for introduction of the human C5aR transgene into the non-human animal in order to achieve homologous recombination. ES cells may be obtained from pre-implantation embryos cultured in vitro and fused with embryos (Evans, M. J., et al. (1981) Nature 292, 154-156; Bradley, M. O., et al. (1984) Nature 309, 255-258; Gossler, et al. (1986) Proc. Natl. Acad. Sci U.S.A. 83, 9065-9069; and Robertson, et al. (1986) Nature 322, 445-448). Transgenes can be efficiently introduced into the ES cells by DNA transfection or by retrovirus-mediated transduction. Such transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The ES cells thereafter colonize the embryo and contribute to the germ line of the resulting chimeric animal. For review see Jaenisch, R. (1988) Science 240, 1468-1474. The transfected embryonal cells may be incubated in vitro for varying amounts of time, or reimplanted into the surrogate host, or both.

Transgenic offspring of the surrogate host may be screened for the presence and/or expression of the transgene by any suitable method. Screening is often accomplished by Southern blot or Northern blot analysis, using a probe that is complementary to at least a portion of the transgene. Western blot analysis using an antibody against the protein encoded by the transgene may be employed as an alternative or additional method for screening for the presence of the transgene product. Typically, in transgenic mice, DNA is prepared from tail tissue and analyzed by Southern analysis or PCR for the transgene. Alternatively, the tissues or cells believed to express the transgene at the highest levels are tested for the presence and expression of the transgene using Southern analysis or PCR, although any tissues or cell types may be used for this analysis. Progeny of the transgenic animals may be obtained by mating the transgenic animal with a suitable partner, or by in vitro fertilization of eggs and/or sperm obtained from the transgenic animal.

Methods of producing transgenic mice via homologous recombination between the endogenous gene and a transgene construct are described by Hanks, M et al (Science 269: 679-682, 1995), which is specifically incorporated herein by reference.

Human C5aR constructs

The introduced polynucleotide may be a wild-type human C5aR sequence, a naturally occurring polymorphism, or a variant or derivative or fragement thereof.

The terms "variant" or "derivative" in relation to the amino acid sequences of the present invention includes any substitution of, variation of, modification of, replacement of, deletion of or addition of one (or more) amino acids from or to the sequence shown in SEQ ID NO:1 providing the resultant amino acid sequence has C5aR activity, preferably having at least 25 to 50% of the activity as the polypeptides presented in the sequence listings, more preferably at least substantially the same activity.

Thus, the human C5aR sequence shown in SEQ ID NO:1 may be modified for use in the present invention. Typically, modifications are made that maintain the native activity of the sequence. Thus, in one embodiment, amino acid substitutions may be made, for example from 1, 2 or 3 to 10, 20 or 30 substitutions provided that the modified sequence retains at least about 25 to 50% of, or substantially the same C5aR signalling. However, in an alternative embodiment, modifications to the amino acid sequences of a polypeptide of the invention may be made intentionally to reduce the biological activity of the polypeptide.

In general, preferably less than 20%, 10% or 5% of the amino acid residues of a variant or derivative are altered as compared with the corresponding region depicted in SEQ ID NO:1.

Conservative substitutions may be made, for example according to the Table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y

In a preferred embodiment of the invention, the C5aR coding sequence is operably linked to a promoter, which may be constitutive or inducible, and other regulatory sequences required for expression in the host animal.

Constructs for use in the present invention include any construct suitable for use in the generation of transgenic animals having the desired levels of expression of the human C5aR-encoding sequence. These constructs may contain cDNA, genomic sequences, or both. Methods for isolating and cloning a desired sequence, as well as suitable constructs for expression of a selected sequence in a host animal, are well known in the art. The construct can include sequences other than the C5aR-encoding sequences. For example, a marker gene such as lac Z may be included in the construct, where upregulation of expression of the encoded sequence will result in an easily detected change in phenotype.

The term "C5aR gene" is used generically to mean a human C5aR gene and isoforms, alternate forms, splice variants, mutated variants, etc. of this human gene. This term is also intended to mean the open reading frame encoding specific polypeptides, introns, and adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 1 kb beyond the coding region, but possibly further in either direction. The DNA sequence encoding C5aR may be cDNA or genomic DNA or a fragment thereof. The gene may be introduced into an appropriate vector for extrachromosomal maintenance or for integration into the host.

The genomic sequences of particular interest comprise the nucleic acid present between the initiation codon and the stop codon, including all of the introns that are normally present in a native chromosome. They may further include the 3' and 5' untranslated

regions found in the mature mRNA. They may further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, etc., including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' or 3' end of the transcribed region. The genomic DNA may be isolated as a fragment of 100 kb or smaller; and substantially free of flanking chromosomal sequence.

The sequences of the 5' regions of the human C5aR gene, and further 5' upstream sequences and 3' downstream sequences, may be utilized for promoter elements, including enhancer binding sites, that provide for expression in tissues where C5aR is normally expressed. The tissue specific expression is useful for providing promoters that mimic the native pattern of expression. Naturally occurring polymorphisms in the promoter region are useful for determining natural variations in expression, particularly those that may be associated with disease. Alternatively, mutations may be introduced into the promoter region to determine the effect of altering expression in experimentally defined systems. Methods for the identification of specific DNA motifs involved in the binding of transcriptional factors are known in the art, e.g sequence similarity to known binding motifs, gel retardation studies, etc. For examples, see Blackwell et al. (1995) Mol Med 1:194-205; Mortlock et al. (1996) Genome Res. 6:327-33; and Joulin and Richard-Foy (1995) Eur J Biochem 232:620-626,

The nucleic acid constructs used in the subject invention may encode all or a part of human C5aR as appropriate. Preferably, the coding sequence for C5aR includes regions sufficient to effect signalling by C5aR. By way of example, regions of the coding sequence for C5aR preferably include segments encoding the amino terminus, extracellular binding regions and/or transmembrane domain regions.

In one embodiment, vectors suitable for use in the present invention may comprise at least one expression control element operably linked to the nucleic acid sequence encoding human C5aR. Expression control elements may be inserted in the vector to control and regulate the expression of the nucleic acid sequence. Examples of expression control elements include, but are not limited to, lac system, operator and promoter regions of phage lamda, yeast promoters, and promoters derived from polyoma, adenovirus, retroviruses, or SV40. The vector may further comprise additional operational elements including, but not limited to, leader sequences, termination codons, polyadenylation signals, and any other sequences necessary or

preferred for the appropriate transcription and/or translation of the nucleic acid sequence encoding human C5aR.

It will be further understood by one skilled in the art that such vectors are constructed using conventional methodology (See e.g. Sambrook et al., (eds.) (1989) "Molecular Cloning, A Laboratory Manual" Cold Spring Harbor Press, Plainview, N.Y.; Ausubel et al., (eds.) (1987) "Current Protocols in Molecular Biology" John Wiley and Sons, New York, N.Y.) or are commercially available.

In some embodiments it may be preferable to express human C5aR in tissues that mimic the native pattern of expression in humans. A specific expression pattern may be accomplished by placing the nucleic acid encoding the human C5aR under the control of an inducible or developmentally regulated promoter, or under the control of a tissue specific or cell type specific promoter. By way of example, specific expression patterns may be accomplished by the use of genomic sequences for human C5aR.

Techniques for *in vitro* mutagenesis of cloned genes are known. Examples of protocols for scanning mutations may be found in Gustin et al., 1993 Biotechniques 14:22; Barany, 1985 Gene 37:111-23; Colicelli et al., 1985 Mol Gen Genet 199:537-9; and Prentki et al., 1984 Gene 29:303-13. Methods for site specific mutagenesis can be found in Sambrook et al., 1989 Molecular Cloning: A Laboratory Manual, CSH Press, pp. 15.3-15.108; Weiner et al., 1993 Gene 126:35-41; Sayers et al., 1992 Biotechniques 13:592-6; Jones and Winistorfer, 1992 Biotechniques 12:528-30; Barton et al., 1990 Nucleic Acids Res 18:7349-55; Marotti and Tomich, 1989 Gene Anal Tech 6:67-70; and Zhu 1989 Anal Biochem 177:120-4.

"Knock outs" and "Knock ins"

Although not necessary to the operability of the invention, the transgenic animals described herein may comprise alterations to endogenous genes in addition to the genetic alterations described above. For example, the host animals may be either "knockouts" and/or "knockins" for a target gene(s) as is consistent with the goals of the invention (e.g, the host animal's endogenous C5aR may be "knocked out" and/or a human C5aR "knocked in"). Knockouts have a partial or complete loss of function in one or both alleles of an endogenous gene of interest (e.g CD4). Knockins have an introduced transgene with altered genetic sequence and/or function from the

endogenous gene. The two may be combined, for example, such that the naturally occurring gene is disabled, and an altered form introduced. For example, it is preferable to knockout the host animal's endogenous C5aR gene, while introducing an a human C5aR gene.

In a knockout, preferably the target gene expression is undetectable or insignificant. For example, a knock-out of an CD4 gene means that function of the CD4 has been substantially decreased so that expression is not detectable or only present at insignificant levels. This may be achieved by a variety of mechanisms, including introduction of a disruption of the coding sequence, e.g. insertion of one or more stop codons, insertion of a DNA fragment, etc., deletion of coding sequence, substitution of stop codons for coding sequence, etc. In some cases the exogenous transgene sequences are ultimately deleted from the genome, leaving a net change to the native sequence. Different approaches may be used to achieve the "knock-out". A chromosomal deletion of all or part of the native gene may be induced, including deletions of the non-coding regions, particularly the promoter region, 3' regulatory sequences, enhancers, or deletions of gene that activate expression of C5aR. A functional knock-out may also be achieved by the introduction of an anti-sense construct that blocks expression of the native genes (for example, see Li and Cohen (1996) Cell 85:319-329). "Knock-outs" also include conditional knock-outs, for example where alteration of the target gene occurs upon exposure of the animal to a substance that promotes target gene alteration, introduction of an enzyme that promotes recombination at the target gene site (e.g. Cre in the Cre-lox system), or other method for directing the target gene alteration postnatally.

A "knockin" of a target gene means an alteration in a host cell genome that results in altered expression or function of a native target gene. Increased (including ectopic) or decreased expression may be achieved by introduction of an additional copy of the target gene, or by operatively inserting a regulatory sequence that provides for enhanced expression of an endogenous copy of the target gene. These changes may be constitutive or conditional, i.e. dependent on the presence of an activator or represser.

Identification of Ligands, Agonists and/or Antagonists of C5aR

Through use of the subject transgenic animals or cells derived therefrom, one can identify ligands or substrates that modulate phenomena associated with C5saR

signalling. Depending on the particular assay, whole transgenic animals of the present invention may be used, or cells derived therefrom. Cells may be freshly isolated from an animal, or may be immortalized in culture. Cells of particular interest are immune cells.

The term "compound" as used herein describes any molecule, e.g. protein, small molecule, polynucleotide, or pharmaceutical, with the capability of preventing or suppressing the molecular and clinical phenomena associated with C5aR signalling.

Candidate compounds encompass numerous chemical classes, though in a preferred embodiment they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate compounds preferably comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate compounds are also found among biomolecules including, but not limited to saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate compounds may be obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, etc. to produce structural analogs.

Screening may also be directed to known pharmacologically active compounds and chemical analogs thereof.

The candidate agent can be administered to the transgenic mammal of the invention (or cells derived from the transgenic mammal) in any manner desired and/or appropriate for delivery of the agent in order to effect a desired result. For example, the candidate agent can be administered by injection (e.g., by injection intravenously, intramuscularly, subcutaneously, or directly into the tissue in which the desired affect is to be achieved), orally, or by any other desirable means. Preferably, the *in vivo* screen will involve a number of animals receiving varying amounts and concentrations of the candidate compound (from no compound to an amount of compound that approaches an upper limit of the amount that can be delivered successfully to the animal), and may include delivery of the agent in different formulation. The compounds can be administered singly or can be combined in combinations of two or more, especially where administration of a combination of agents may result in a synergistic effect. The effect of agent administration upon the transgenic rodent can be monitored by conventional methodology.

The range of candidate compounds contemplated herein include C5aR inhibitory compounds or antagonists of a biological function of C5aR. In one embodiment, the compound is selected from the group consisting of: a peptide, including a peptide derived from C5aR and capable of inhibiting, reducing or repressing a C5aR function, a C5aR dominant-negative mutant; a non-C5aR peptide inhibitor of C5aR; an antibody or antibody fragment which binds to C5aR and inhibits a C5aR function; a small organic molecule, and nucleic acid, including nucleic acid encoding said peptide derived from C5aR or said non-C5aR peptide inhibitor, an antisense nucleic acid directed against C5aR-encoding mRNA, or an anti-C5aR ribozyme, or a small interfering RNA (RNAi) that targets C5aR gene expression. A number of these types of compound are discussed below.

Small molecules

Candidate compounds encompass numerous chemical classes, though preferably they are organic molecules, preferably small organic compounds having a molecular weight of more than 100 and less than about 2,500 daltons. Candidate compounds comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemicals groups. The candidate compounds often comprise cyclical carbon or heterocyclic structures and/or

aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate compounds are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines and derivatives, structural analogs or combinations thereof. In fact, virtually any small organic molecule that is potentially capable of binding to a biological target molecule of interest may find use in the present invention provided that it is sufficiently soluble and stable in aqueous solutions to be tested for its ability to bind to the biological target molecule.

Candidate compounds are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means. Known pharmacological compounds may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification to produce structural analogs.

Protein or Peptide inhibitors

In another embodiment, the candidate compounds are proteins. By "*protein*" in this context it is meant at least two covalently attached amino acids, which includes proteins, polypeptides, oligopeptides and peptides. The protein may be made up of naturally occurring amino acids and peptide bonds, or synthetic peptidomimetic structures. Thus "*amino acid*", or "*peptide residue*", as used herein means both naturally occurring and synthetic amino acids. For example, homophenylalanine, citrulline and noreleucine are considered amino acids for the purposes of the invention. "*Amino acid*" also includes imino acid residues such as proline and hydroxyproline. The side chains may be in either the (R) or the (S) configuration. In the preferred embodiment, the amino acids are in the (S) or (L)-configuration. If non-naturally occurring side chains are used, non-amino acid substituents may be used, for example to prevent or retard *in vivo* degradations.

In a further preferred embodiment, the candidate compounds are naturally occurring proteins or fragments of naturally occurring proteins. Thus, for example, cellular extracts containing proteins, or random or directed digests of proteinaceous cellular extracts, may be used. In this way libraries of prokaryotic and eukaryotic proteins may be made. Particularly preferred in this embodiment are libraries of bacterial, fungal, viral, and mammalian proteins, with the latter being preferred, and human proteins being especially preferred.

In a further preferred embodiment, the candidate compounds are peptides of from about 5 to about 30 amino acids, with from about 5 to about 20 amino acids being preferred, and from about 7 to about 15 being particularly preferred. The peptides may be digests of naturally occurring proteins, random peptides, or "*biased*" random peptides. By "*randomized*" or grammatical equivalents herein is meant that each peptide consists of essentially random amino acids. Since generally these random peptides are chemically synthesized, they may incorporate any amino acid at any position. The synthetic process can be designed to generate randomized proteins to allow the formation of all or most of the possible combinations over the length of the sequence, thus forming a library of randomized candidate bioactive proteinaceous compounds.

Preparation and screening of combinatorial chemical libraries are well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Pat. No. 5,010,175, Furka (1991) Int. J. Pept. Prot. Res., 37: 487-493, Houghton *et al.* (1991) Nature, 354: 84-88). Peptide synthesis is by no means the only approach envisioned and intended for use with the present invention. Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (PCT Publication No WO 91/19735, Dec. 26, 1991), encoded peptides (PCT Publication WO 93/20242, Oct. 14, 1993), random bio-oligomers (PCT Publication WO 92/00091, Jan. 9, 1992), benzodiazepines (U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al.*, (1993) Proc. Nat. Acad. Sci. USA 90: 69096913), vinylogous polypeptides (Hagihara *et al.* (1992) J. Amer. Chem. Soc. 114: 6568), nonpeptidal peptidomimetics with a Beta D Glucose scaffolding (Hirschmann *et al.*, (1992) J. Amer. Chem. Soc. 114: 92179218), analogous organic syntheses of small compound libraries (Chen *et al.* (1994) J. Amer. Chem. Soc. 116: 2661), oligocarbamates (Cho, *et al.*, (1993) Science 261:1303), and/or peptidyl phosphonates (Campbell *et al.*, (1994) J. Org. Chem. 59: 658). See, generally, Gordon *et al.*, (1994) J.

Med. Chem. 37:1385, nucleic acid libraries, peptide nucleic acid libraries (see, e.g., U.S. Pat. No. 5,539,083) antibody libraries (see, e.g., Vaughn *et al.* (1996) Nature Biotechnology, 14(3): 309-314), and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang *et al.* (1996) Science, 274: 1520-1522, and U.S. Pat. No. 5,593,853), and small organic molecule libraries (see, e.g., benzodiazepines, Baum (1993) C&EN, Jan 18, page 33, isoprenoids U.S. Pat. No. 5,569,588, thiazolidinones and metathiazanones U.S. Pat. No. 5,549,974, pyrrolidines U.S. Pat. Nos. 5,525,735 and 5,519,134, morpholino compounds U.S. Pat. No. 5,506,337, benzodiazepines U.S. Pat. No. 5,288,514, and the like).

Devices for the preparation of combinatorial libraries are commercially available (see, e.g., 357 MPS, 390 MPS, Advanced Chem Tech, Louisville Ky., Symphony, Rainin, Woburn, Mass., 433A Applied Biosystems, Foster City, Calif., 9050 Plus, Millipore, Bedford, Mass.).

In one embodiment, peptidyl C5aR inhibitors are chemically or recombinantly synthesized as oligopeptides (usually 10-25 amino acids in length) derived from the C5aR sequence (SEQ ID NO:2 or 4). Alternatively, C5aR fragments are produced by digestion of native or recombinantly produced C5aR by, for example, using a protease, e.g., trypsin, thermolysin, chymotrypsin, or pepsin. Computer analysis (using commercially available software, e.g. MacVector, Omega, PCGene, Molecular Simulation, Inc.) is used to identify proteolytic cleavage sites. The proteolytic or synthetic fragments can comprise as many amino acid residues as are necessary to partially or completely inhibit C5aR function. Preferred fragments will comprise at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more amino acids in length.

Protein or peptide inhibitors may also be dominant-negative mutants of C5aR. The term "dominant-negative mutant" refers to a C5aR polypeptide that has been mutated from its natural state and that interacts with a protein that C5aR normally interacts with thereby preventing endogenous native C5aR from forming the interaction.

Anti-C5a and Anti-C5aR Antibodies

The term "antibody" as used in this invention includes intact molecules as well as fragments thereof, such as Fab, F(ab')2, and Fv which are capable of binding an

epitopic determinant of C5aR. These antibody fragments retain some ability to selectively bind with its antigen and are defined as follows:

(1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule can be produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain;

(2) Fab', the fragment of an antibody molecule can be obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule;

(3) (Fab')2, the fragment of the antibody that can be obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; F(ab)2 is a dimer of two Fab' fragments held together by two disulfide bonds;

(4) Fv, defined as a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and

(5) Single chain antibody ("SCA"), defined as a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule.

Methods of making these fragments are known in the art. (See for example, Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York (1988), incorporated herein by reference).

Antibodies of the present invention can be prepared using intact C5a or C5aR or fragments thereof as the immunizing antigen. A peptide used to immunize an animal can be derived from translated cDNA or chemical synthesis and is purified and conjugated to a carrier protein, if desired. Such commonly used carriers which are chemically coupled to the peptide include keyhole limpet hemocyanin (KLH), thyroglobulin, bovine serum albumin (BSA), and tetanus toxoid. The coupled peptide may then be used to immunize the animal (e.g., a mouse or a rabbit).

If desired, polyclonal antibodies can be further purified, for example, by binding to and elution from a matrix to which the peptide to which the antibodies were raised is bound. Those of skill in the art will know of various techniques common in the immunology arts for purification and/or concentration of polyclonal antibodies, as well as monoclonal antibodies (See for example, Coligan, et al., Unit 9, Current Protocols in Immunology, Wiley Interscience, 1991, incorporated by reference).

Monoclonal antibodies may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture, such as, for example, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler et al. Nature 256, 495-497, 1975; Kozbor et al., J. Immunol. Methods 81, 31-42, 1985; Cote et al., Proc. Natl. Acad. Sci. USA 80, 2026-2030, 1983; Cole et al., Mol. Cell Biol. 62, 109-120, 1984).

Methods known in the art allow antibodies exhibiting binding for C5a or C5aR to be identified and isolated from antibody expression libraries. For example, a method for the identification and isolation of an antibody binding domain which exhibits binding to C5aR is the bacterio-phage a vector system. This vector system has been used to express a combinatorial library of Fab fragments from the mouse antibody repertoire in Escherichia coli (Huse, et al., Science, 246:1275-1281, 1989) and from the human antibody repertoire (Mullinax, et al., Proc. Nat. Acad. Sci., 87:8095-8099, 1990). This methodology can also be applied to hybridoma cell lines expressing monoclonal antibodies with binding for a preselected ligand. Hybridomas which secrete a desired monoclonal antibody can be produced in various ways using techniques well understood by those having ordinary skill in the art and will not be repeated here. Details of these techniques are described in such references as Monoclonal Antibodies-Hybridomas: A New Dimension in Biological Analysis, Edited by Roger H. Kennett, et al., Plenum Press, 1980; and U.S. 4,172,124, incorporated by reference.

In addition, methods of producing chimeric antibody molecules with various combinations of "humanized" antibodies are known in the art and include combining murine variable regions with human constant regions (Cabily, et al. Proc. Natl. Acad. Sci. USA, 81:3273, 1984), or by grafting the murine-antibody complementarity determining regions (CDRs) onto the human framework (Riechmann, et al., Nature 332:323, 1988).

Antisense compounds

The term "antisense compounds" encompasses DNA or RNA molecules that are complementary to at least a portion of a target mRNA molecule (Izant and Weintraub, 1984; Izant and Weintraub, 1985) and capable of interfering with a post-transcriptional event such as mRNA translation. Antisense oligomers complementary to at least about 15 contiguous nucleotides of the target-encoding mRNA are preferred, since they are easily synthesized and are less likely to cause problems than larger molecules when introduced into the target mRNA producing cell. The use of antisense methods is well known in the art (Marcus-Sakura, 1988).

Catalytic RNA molecules

The term catalytic RNA refers to an RNA or RNA-containing molecule (also known as a "ribozyme") which specifically recognizes a distinct substrate and catalyzes the chemical modification of this substrate. The nucleic acid bases in the catalytic nucleic acid can be bases A, C, G, T and U, as well as derivatives thereof. Derivatives of these bases are well known in the art.

Typically, the catalytic nucleic acid contains an antisense sequence for specific recognition of a target nucleic acid, and a nucleic acid cleaving enzymatic activity (also referred to herein as the "catalytic domain").

The types of ribozymes that are particularly useful in this invention are the hammerhead ribozyme (Haseloff and Gerlach 1988, Perriman et al, 1992) and the hairpin ribozyme (Shippy et al, 1999).

The ribozymes used in this invention can be chemically synthesized using methods well known in the art. The ribozymes can also be prepared from a DNA molecule (that upon transcription, yields an RNA molecule) operably linked to an RNA polymerase promoter, e.g., the promoter for T7 RNA polymerase or SP6 RNA polymerase. When the vector also contains an RNA polymerase promoter operably linked to the DNA molecule, the ribozyme can be produced in vitro upon incubation with RNA polymerase and nucleotides. In a separate embodiment, the DNA can be inserted into an expression cassette or transcription cassette.

dsRNA

dsRNA is particularly useful for specifically inhibiting the production of a particular protein. Although not wishing to be limited by theory, Dougherty and Parks (1995) have provided a model for the mechanism by which dsRNA can be used to reduce protein production. This model was modified and expanded by Waterhouse et al (1998). This technology relies on the presence of dsRNA molecules that contain a sequence that is essentially identical to the mRNA of the gene of interest. Conveniently, the dsRNA can be produced in a single open reading frame in a recombinant vector or host cell, where the sense and anti-sense sequences are flanked by an unrelated sequence which enables the sense and anti-sense sequences to hybridize to form the dsRNA molecule with the unrelated sequence forming a loop structure. The design and production of suitable dsRNA molecules targeted against genes of interest is well within the capacity of a person skilled in the art, particularly considering Dougherty and Parks (1995), Waterhouse et al (1998), WO 99/32619, WO 99/53050, WO 99/49029, and WO 01/34815.

As used herein, the terms "small interfering RNA", and "RNAi" refer to homologous double stranded RNA (dsRNA) that specifically targets a gene product, thereby resulting in a null or hypomorphic phenotype. Specifically, the dsRNA comprises two nucleotide sequences derived from the target RNA and having self-complementarity such that they can anneal, and interfere with expression of a target gene, presumably at the post-transcriptional level. RNAi molecules are described by Fire et al (1998) and reviewed by Sharp (1999).

Phenotypes associated with C5aR signalling

In the methods of the present invention, a putative compound is administered to the transgenic animal and a response of the transgenic animal to the putative compound is measured. Preferably, the response of the transgenic mammal is compared to the response of a "normal" or wild-type mouse or, alternatively, compared to a transgenic animal control (without administration of the compound).

Accordingly, in one aspect the present invention provides a method of identifying a compound that modulates C5aR activity, the method comprising (i) administering a candidate compound to a transgenic mammal of the present invention under conditions

in which at least one phenotype associated with C5aR signalling is expressed; and (ii) monitoring development of the phenotype following administration of the compound.

The phenotype monitored may be any indicator of C5aR signalling, including the following:

(a) inflammatory or allergic diseases and conditions, including respiratory allergic diseases such as asthma, allergic rhinitis, hypersensitivity lung diseases, hypersensitivity pneumonitis, interstitial lung diseases (ILD) (e.g., idiopathic pulmonary fibrosis, or ILD associated with rheumatoid arthritis, systemic lupus erythematosus, ankylosing spondylitis, systemic sclerosis, Sjogren's syndrome, polymyositis or dermatomyositis); anaphylaxis or hypersensitivity responses, drug allergies (e.g., to penicillin, cephalosporins), insect sting allergies; inflammatory bowel diseases, such as Crohn's disease and ulcerative colitis; spondyloarthropathies; scleroderma; psoriasis and inflammatory dermatoses such as dermatitis, eczema, atopic dermatitis, allergic contact dermatitis, urticaria; vasculitis (e.g., necrotizing, cutaneous, and hypersensitivity vasculitis);

(b) autoimmune diseases, such as arthritis (e.g., rheumatoid arthritis, psoriatic arthritis), multiple sclerosis, systemic lupus erythematosus, myasthenia gravis, juvenile onset diabetes, nephritides such as glomerulonephritis, autoimmune thyroiditis, Behcet's disease;

(c) graft rejection (e.g., in transplantation), including allograft rejection or graft-versus-host disease;

(d) atherosclerosis;

(e) cancers with leukocyte infiltration of the skin or organs;

(f) diseases or conditions (including C5aR-mediated diseases or conditions), in which undesirable inflammatory responses are to be inhibited can be treated, including, but not limited to, reperfusion injury, stroke, adult respiratory distress syndrome, certain hematologic malignancies, cytokine-induced toxicity (e.g., septic shock, endotoxic shock), polymyositis, dermatomyositis, pemphigoid, Alzheimers Disease and granulomatous diseases including sarcoidosis.

Other phenotypes include immunosuppression, such as that in individuals with immunodeficiency syndromes such as AIDS, individuals undergoing radiation therapy, chemotherapy, therapy for autoimmune disease or other drug therapy (e.g., corticosteroid therapy), which causes immunosuppression; and immunosuppression due congenital deficiency in receptor function or other causes.

A number of *in vivo* models of inflammation are available and can be used to induce suitable C5aR associated phenotypes in transgenic mammals of the present invention. For example, rheumatoid arthritis can be evaluated using an animal (e.g. mouse) model of collagen-induced arthritis (Trentham et al (1977) *J Exp Med* 146: 857-868), K/BxN serum-induced arthritis (Kouskoff *et al* (1996) Cell 87:811-822), antigen-induced arthritis, or adjuvant-induced arthritis (Pearson CM (1956) Proc Soc. Expe Biblo Med 91:95-101).

Further examples of suitable animal models include: cecal ligation puncture (CLP) model of sepsis (Huber-Lang, M. S., *et al.* (2002) Faseb J 16(12): 1567-74); rat model of RA (Woodruff, T. M., et al. (2002) *Arthritis Rheum* **46**(9): 2476-85); porcine model of sepsis (Mohr, M., *et al.* (1998) *Eur J Clin Invest* 28(3): 227-34); immune complex-induced lung disease; pancreatitis associated lung injury (Bhatia, M., *et al.* (2001) *Am J Physiol Gastrointest Liver Physiol* 280(5): G974-8); acute lung injury; renal ischaemia-reperfusion injury; collagen-induced arthritis; and experimental airway disease (asthma like model).

In another example, leukocyte infiltration upon intradermal injection of a candidate compound can be monitored (see e.g., Van Damme, J. *et al.*, J. Exp. Med., 176: 59-65 (1992); Zachariae, C. O. C. *et al.*, J. Exp. Med. 171: 2177-2182 (1990); Jose, P. J. et al., J. Exp. Med. 179: 881-887 (1994)). In one embodiment, skin biopsies are assessed histologically for infiltration of leukocytes (e.g., eosinophils, granulocytes). A decrease of the extent of infiltration in the presence of the candidate compound as compared with the extent of infiltration in the absence of the candidate compound is indicative of inhibition of C5aR signalling.

Examples of preferred phenotypes are discussed briefly below.

Sepsis

Sepsis is a severe illness caused by overwheming infection of the bloodstream by toxin-producing bacteria. It is has been recognized that the innate immune system can be perturbed during sepsis, as evidenced by extensive activation of the inflammatory, complement, and clotting systems, together with the appearance in plasma of cytokines and chemokines. This "cytokine storm" as it has been termed unleashes numerous inflammatory mediators that contribute to multiple-organ dysfunction or failure and death.

Recent reviews of sepsis reflect on past attempts at finding treatments for this disease and suggest future directions (Crowther and Marshall (2001) Jama 286(15): 1894-6; Cohen, J. (2002) Nature 420(6917):885-91; Cross and Opal (2003) Ann Intern Med 138(6): 502-5).

Hotchkiss and Karl (N Engl J Med 348(2):138-50, 2003) speculate that the efficacy shown by the anti-coagulant recombinant activated protein C in reducing mortality in sepsis may be due in part to its anti-inflammatory effects. Some of the effects (direct and indirect) of activated protein C are blocking production of cytokines, blocking cell adhesion, inhibiting neutrophil recruitment, mast cell degranulation and platelet activation. C5a mediates leukocyte chemotaxis, release of histamine from mast cells, enhancement of neutrophil-endothelial cell adhesion and induction of cytokine production through binding to the C5a receptor.

During sepsis in humans, unregulated activation of the complement system results in excessive generation of anaphylatoxins, especially C3a and C5a and ensuing dysfunction of neutrophils. In later stages of sepsis, neutrophils have suppressed chemotactic responsiveness, depressed enzyme release, alterations of intracellular pH and a defective respiratory burst (diminished production of reactive oxygen species, especially H_2O_2), leading to impaired bactericidal activity. These defects have been shown to be C5a dependent. During experimental sepsis blood neutrophils have a greatly diminished ability to bind C5a, have impaired chemotactic responses to C5a and a loss of H_2O_2 production. These defects can be prevented in CLP induced sepsis by treating animals with anti-C5a antibodies. This treatment reduced bacteremia and greatly improved survival. This indicates that sepsis induces excessive generation of C5a, which, in turn, leads to serious functional defects in neutrophils.

Antibodies to C5aR can also protect against death from sepsis in animal models. It has been shown that C5aR immunoreactivity and mRNA expression both are increased in epithelial cells of lung, kidney, liver, heart and thymus early in experimental sepsis. Mice injected at the start of CLP with antibodies to C5aR showed dramatically improved survival when compared with animals receiving nonspecific IgG. In anti-C5aR-treated mice, serum levels of IL-6 and TNF-α and bacterial counts in various organs were significantly reduced during CLP when compared with control CLP animals. These studies demonstrated that blockade of C5aR is highly protective from the lethal outcome of sepsis.

It has been suggested that it is the location of C5a that may be critical in terms of its potential to protect or harm. Local generation of C5a in tissues is essential for early control of infection. A C5a gradient is established causing leukocyte chemotaxis. At higher C5a concentrations chemotaxis is arrested and the cells produce their toxic oxygen burst. In sepsis, diffuse complement activation in the blood leads to an excess of diffuse intravascular C5a which cripples neutrophils allowing unrestrained proliferation of bacteria. Simultaneously sequestration of the neutrophils in the microcirculation injures the lung, kidney and liver. In this model the C5aR on leukocytes rather than the receptor on the epithelial cells of the liver, lung etc. may be critical. Intervention, by way of C5aR antagonists may prove therapeutically valuable.

Ischaemia - Reperfusion Injury (IR)

The complement system is an important mediator of inflammatory tissue damage in various diseases including ischaemia-reperfusion (IR) injury.

The relative roles of C5a and the membrane attack complex C5b-9 in mediating IR injury seem to vary. In some models C5a is the important mediator. Antagonists of C5aR protect mice and rats against IR injury in small intestine and kidney (Heller et al. (1999) J Immunol 163(2): 985-94; Arumugam et al. (2003) Kidney Int 63(1): 134-42; Arumugam et al. (2002) J Surg Res 103(2):260-7). In a porcine myocardial IR injury model a C5aR antagonist reduced infarct size markedly (Riley et al. (2000) J Thorac Cardiovasc Surg **120**(2): 350-8).

Studies with antibodies to C5 also demonstrate similar reductions in IR injury (Fitch et al. (1999) Circulation 100(25):2499-506; de Vries et al (2003) Transplantation 75(3): 375-82).

Clinical and experimental studies have shown that IR of organs such as the kidney, liver, lungs and heart induces a rapid release of various cytokines. While many of these molecules decrease in level after reperfusion, IL-8 significantly increases. IL-8 is a potent neutrophil chemoattractant. Levels of IL-8 negatively correlate with lung function and high levels are associated with increased risk of death in lung transplantation. Intravenous administration of anti-IL-8 antibodies at the beginning of reperfusion markedly reduces lung injury and neutrophil infiltration (de Perrot et al (2003) Am J Respir Crit Care Med 167(4):490-511).

Evidence shows that IR injury occurs in a biphasic pattern. Macrophages activated during ischaemia mediate the early phase of injury whereas neutrophils and lymphocytes are primarily involved in the second, delayed phase. The recruitment of neutrophils and lymphocytes results from the release of cytokines before and after reperfusion.

Since C5a is a potent chemoattractant for various myeloid cells including macrophages and neutrophils and also induces cytokine production it is envisaged that blocking C5aR with specific antibodies could be beneficial in preventing IL-8 release and neutrophil migration and thus reduce IR injury.

Acute Respiratory Distress Syndrome (ARDS) & Acute Lung Injury (ALI)

ARDS and ALI are syndromes that result from pulmonary edema and inflammation. The development of ARDS/ALI is associated with several clinical disorders including pulmonary injury from pneumonia infection, aspiration of gastric contents, trauma, sepsis, acute pancreatitis, drug overdose and cardiopulmonary bypass. Sepsis is the leading cause of ARDS/ALI. As with IR, the inflammatory response in ALI is associated with recruitment of large numbers of neutrophils and monocytes into the distal airspaces of the lung. Proinflammatory molecules such as cytokines, oxygen radicals and proteases play a role and excessive inflammation may worsen ARDS/ALI (Ware and Matthay (2000) N Engl J Med 342(18): 1334-49; Brower et al. (2001) Chest 120(4): 1347-67).

It has been suggested that anti-inflammatory strategies to reduce the number of neutrophils migrating into extravascular spaces in the lung, to reduce neutrophil adhesion to the lung endothelium, and to reduce release of chemotactic factors could be beneficial (Brower et al. (2001) Chest 120(4): 1347-67)..

Studies of patients with ARDS reveals that high levels of IL-8 are present in the BAL fluid or pulmonary edema fluid in the early phases of the disease. Furthermore, antibodies that neutralise IL-8 reduced lung injury in a rabbit model (Brower et al. (2001) Chest 120(4): 1347-67).

For ARDS and ALI blocking C5aR may also be a therapeutically viable approach. C5a is a potent chemotactic molecule and stimulator of cytokine release. Studies described above which used molecules that antagonise C5a or C5aR demonstrated reduction in injury in various models of inflammation.

Production of compounds identified in the screening methods of the invention

In one embodiment of the invention, the subject method further comprises producing or synthesizing the compound that is tested on the genetically modified animal.

Peptidyl compounds are conveniently made by standard peptide synthesis, such as the Merrifield method of synthesis (Merrifield, *J Am Chem Soc, 85*,:2149-2154, 1963) and the myriad of available improvements on that technology (see e.g., Synthetic Peptides: A User's Guide, Grant, ed. (1992) W.H. Freeman & Co., New York, pp. 382; Jones (1994) The Chemical Synthesis of Peptides, Clarendon Press, Oxford, pp. 230.); Barany, G. and Merrifield, R.B. (1979) in *The Peptides* (Gross, E. and Meienhofer, J. eds.), vol. 2, pp. 1-284, Academic Press, New York; Wünsch, E., ed. (1974) *Synthese von Peptiden in Houben-Weyls Metoden der Organischen Chemie* (Müler, E., ed.), vol. 15, 4th edn., Parts 1 and 2, Thieme, Stuttgart; Bodanszky, M. (1984) *Principles of Peptide Synthesis*, Springer-Verlag, Heidelberg; Bodanszky, M. & Bodanszky, A. (1984) *The Practice of Peptide Synthesis, Springer-Verlag,* Heidelberg; Bodanszky, M. (1985) *Int. J. Peptide Protein Res.* 25, 449-474.

Preferably, the peptide is synthesized on a solid phase support, such as, for example, a polystyrene gel bead comprising polystyrene cross-linked with divinylbenzene,

preferably 1% (w.w) divinylbenzene, which is further swollen using lipophilic solvent, such as, for example dichloromethane or dimethylformamide (DMF). The polystyrene can be functionalized by addition of chloromethane or amino methyl groups. Alternatively, cross-linked and functionalized polydimethyl-acrylamide gel can be used once swollen and solvated using DMF or dipolar aprotic solvent. Other solid phase supports known to those skilled in the art can also be used for peptide synthesis, such as, for example, polyethylene glycol-derived bead produced by grafting polyethylene glycol to the surface of inert polystyrene beads. Preferred commercially available solid phase supports include PAL-PEG-PS, PAC-PEG-PS, KA, KR, or TGR (Applied Biosystems, CA 94404, USA).

For solid phase peptide synthesis, blocking groups that are stable to the repeated treatments necessary for removal of the amino blocking group of the growing peptide chain and for repeated amino acid couplings, are used for protecting the amino acid side-chains during synthesis and for masking undesired reactivity of the α-amino, carboxyl or side chain functional groups. Blocking groups (also called protecting groups or masking groups) thus protect the amino group of the amino acid having an activated carboxyl group that is involved in the coupling reaction, or protect the carboxyl group of the amino acid having an acylated amino group that is involved in the coupling reaction.

During synthesis, coupling occurs following removal of a blocking group without the disruption of a peptide bond, or any protecting group attached to another part of the peptide. Additionally, the peptide-resin anchorage that protects the C-terminus of the peptide is protected throughout the synthetic process until cleavage from the resin is required. Accordingly, by the judicious selection of orthogonally protected α-amino acids, amino acids are added at desired locations to a growing peptide whilst it is still attached to the resin.

Preferred amino blocking groups are easily removable but sufficiently stable to survive conditions for the coupling reaction and other manipulations, such as, for example, modifications to the side-chain groups. In one embodiment, amino blocking groups are selected from the group consisting of: (i) a benzyloxycarbonyl group (Z or carbocenzoxy) that is removed easily by catalytic hydrogenation at room temperature and ordinary pressure, or using sodium in liquid ammonia and hydrobromic acid in acetic acid; (ii) a urethane derivative; (iii) a t-Butoxycarbonyl group (Boc) that is

introduced using t-butoxycarbonyl azide or di-tert-butyldicarbonate and removed using mild acid such as, for example, trifluoroacetic acid (50% TFA in dichloromethane), or HCl in acetic acid/dioxane/ethylacetate; (iv) a 9-fluorenylmethyloxycarbonyl group (Fmoc) that is cleaved under mildly basic, non-hydrolytic conditions, such as, for example, using a primary or secondary amine (eg. 20% piperidine in dimethyl formamide); (v) a 2-(4-biphenylyl) propyl(2)oxycarbonyl group (Bpoc); (vi) a 2-nitro-phenylsulfenyl group (Nps); and (vii) a dithia-succionyl group (Dts). Boc is widely used to protect the N-terminus in Fmoc chemistry, or Fmoc is widely used to protect the N-terminus in Boc chemistry.

Side chain-protecting groups will vary for the functional side chains of the amino acids forming the peptide being synthesized. Side-chain protecting groups are generally based on the Bzl group or the tBu group. Amino acids having alcohols or carboxylic acids in the side-chain are protected as Bzl ethers, Bzl esters, cHex esters, tBu ethers, or tBu esters. Side-chain protection of Fmoc amino acids requires blocking groups that are ideally base stable and weak acid (TFA) labile. Many different protecting groups for peptide synthesis have been described (*see* The Peptides, Gross *et al.* eds., Vol. 3, Academic Press, New York, 1981). For example, the 4-methoxy-2,3,6-trimethylphenylsulfonyl (Nd- Mtr) group is useful for Arginine side-chain protection, however deprotection of Arg(Mtr) requires prolonged TFA treatment. A number of soft acid (TFA, thalium (III) trifluoroacetate/TFA) labile groups, or TFA stable but thalium (III) trifluoroacetate/TFA labile groups, or soft acid stable groups are used to protect Cystine.

The two most widely used protection strategies are the Boc/Bzl- and the Fmoc/tBu-strategies. In Boc/Bzl, Boc is used for amino protection and the side-chains of the various amino acids are protected using Bzl- or cHex-based protecting groups. A Boc group is stable under catalytic hydrogenation conditions and is used orthogonally along with a Z group for protection of many side chain groups. In Fmoc/tBu, Fmoc is used for amino protection and the side-chains are protected with tBu-based protecting groups.

Alternatively, the peptidyl compound is produced by the recombinant expression of nucleic acid encoding the amino acid sequence of said peptide. Random peptide-encoding libraries are particularly preferred for such purposes, because they provide a wide range of different compounds to test. Alternatively, naturally-occurring nucleic

acids can be screened. According to this embodiment, nucleic acid encoding the peptidyl compound is produced by standard oligonucleotide synthesis or derived from a natural source and cloned into a suitable expression vector in operable connection with a promoter or other regulatory sequence capable of regulating expression in a cell-free system or cellular system..

Oligonucleotides are preferably synthesized with linker or adaptor sequences at the 5'- and 3'-ends to facilitate subsequent cloning into a suitable vector system using standard techniques.

Placing a nucleic acid molecule under the regulatory control of, i.e., "in operable connection with", a promoter sequence means positioning said molecule such that expression is controlled by the promoter sequence, generally by positioning the promoter 5' (upstream) of the peptide-encoding sequence.

The prerequisite for producing intact peptides in bacteria such as *E. coli* is the use of a strong promoter with an effective ribosome binding site. Typical promoters suitable for expression in bacterial cells such as *E. coli* include, but are not limited to, the *lacz* promoter, temperature-sensitive λ_L or λ_R promoters, T7 promoter or the IPTG-inducible *tac* promoter. A number of other vector systems for expressing the nucleic acid molecule of the invention in *E. coli* are well-known in the art and are described, for example, in Ausubel *et al* (*In:* Current Protocols in Molecular Biology. Wiley Interscience, ISBN 047150338, 1987) or Sambrook *et al* (*In:* Molecular cloning, A laboratory manual, second edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989). Numerous plasmids with suitable promoter sequences for expression in bacteria and efficient ribosome binding sites have been described, such as for example, pKC30 (λ_L: Shimatake and Rosenberg, *Nature 292*, 128, 1981); pKK173-3 (*tac*: Amann and Brosius, *Gene 40,* 183, 1985), pET-3 (T7: Studier and Moffat, *J. Mol. Biol. 189,* 113, 1986); the pBAD/TOPO or pBAD/Thio-TOPO series of vectors containing an arabinose-inducible promoter (Invitrogen, Carlsbad, CA), the latter of which is designed to also produce fusion proteins with thioredoxin to enhance solubility of the expressed protein; the pFLEX series of expression vectors (Pfizer Inc., CT, USA); or the pQE series of expression vectors (Qiagen, CA), amongst others.

Typical promoters suitable for expression in viruses of eukaryotic cells and eukaryotic cells include the SV40 late promoter, SV40 early promoter and cytomegalovirus

(CMV) promoter, CMV IE (cytomegalovirus immediate early) promoter amongst others. Preferred vectors for expression in mammalian cells (eg. 293, COS, CHO, 10T cells, 293T cells) include, but are not limited to, the pcDNA vector suite supplied by Invitrogen, in particular pcDNA 3.1 myc-His-tag comprising the CMV promoter and encoding a C-terminal 6xHis and MYC tag; and the retrovirus vector pSRαtkneo (Muller *et al., Mol. Cell. Biol., 11*, 1785, 1991). The vector pcDNA 3.1 myc-His (Invitrogen) is particularly preferred for expressing peptides in a secreted form in 293T cells, wherein the expressed peptide or protein can be purified free of conspecific proteins, using standard affinity techniques that employ a Nickel column to bind the protein via the His tag.

A wide range of additional host/vector systems suitable for expressing peptides are available publicly, and described, for example, in Sambrook *et al* (*In:* Molecular cloning, A laboratory manual, second edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989).

Means for introducing the nucleic acid or a gene construct comprising same into a cell for expression are well-known to those skilled in the art. The technique used for a given organism depends on the known successful techniques. Means for introducing recombinant DNA into animal cells include microinjection, transfection mediated by DEAE-dextran, transfection mediated by liposomes such as by using lipofectamine (Gibco, MD, USA) and/or cellfectin (Gibco, MD, USA), PEG-mediated DNA uptake, electroporation and microparticle bombardment such as by using DNA-coated tungsten or gold particles (Agracetus Inc., WI, USA) amongst others.

Techniques for synthesizing small organic compounds will vary considerably depending upon the compound, however such methods will be well known to those skilled in the art. In one embodiment, informatics is used to select suitable chemical building blocks from known compounds, for producing a combinatorial library. For example, QSAR(Quantitative Structure Activity Relationship) modelling approach uses linear regressions or regression trees of compound structures to determine suitability. The software of the Chemical Computing Group, Inc.(Montreal, Canada) uses high-throughput screening experimental data on active as well as inactive compounds, to create a probabilistic QSAR model, which is subsequently used to select lead compounds. The Binary QSAR method is based upon three characteristic properties of compounds that form a "descriptor" of the likelihood that a particular compound will or

will not perform a required function: partial charge, molar refractivity (bonding interactions), and logP (lipophilicity of molecule). Each atom has a surface area in the molecule and it has these three properties associated with it. All atoms of a compound having a partial charge in a certain range are determined and the surface areas (Van der Walls Surface Area descriptor) are summed. The binary QSAR models are then used to make activity models or ADMET models, which are used to build a combinatorial library. Accordingly, information from known appetite suppressants and non-suppressants, including lead compounds identified in initial screens, can be used to expand the list of compounds being screened to thereby identify highly active compounds.

In one embodiment, the subject method further comprises formulating the identified compound for administration to a non-human animal or a human. The formulations can be suitable for administration by injection by a subcutaneous, intravenous, intranasal, or intraperitoneal route. Alternatively, they can be suitable for oral administration in the form of feed additives, tablets, troches, etc.

The compounds are conveniently formulated in a suitable excipient or diluent, such as, for example, an aqueous solvent, non-aqueous solvent, non-toxic excipient, such as a salt, preservative, buffer and the like. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oil and injectable organic esters such as ethyloleate. Aqueous solvents include water, alcoholic/aqueous solutions, saline solutions, parenteral vehicles such as sodium chloride, Ringer's dextrose, etc. Preservatives include antimicrobial, anti-oxidants, chelating agents and inert gases. The pH and exact concentration of the various components the formulation suitable for administration to the animal are adjusted according to routine skills in the art. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic. Pharmaceutically compatible binding agents, and/or adjuvant materials can be

included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring. For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

Optionally, the formulation will also include a carrier, such as, for example, bovine serum albumin (BSA), keyhole limpet hemocyanin (KLH), ovalbumin, mouse serum albumin, rabbit serum albumin and the like. Means for conjugating peptides to carrier proteins are also well known in the art and include glutaraldehyde, m-maleimidobencoyl-N-hydroxysuccinimide ester, carbodiimide and bis-biazotized benzidine.

The present invention will now be illustrated by the following Examples, which are not intended to be limiting in any way. The teachings of all references cited herein are incorporated herein by reference.

Example 1: Generation of targeting construct for producing human C5aR knock-in mouse

To produce a mutant mouse strain by homologous recombination, two major elements are needed. An embryonic stem (ES) cell line capable of contributing to the germ line, and a targeting construct containing target-gene sequences with the desired mutation. Maintaining ES cells in their undifferentiated is accomplished by growing cells on a layer of feeder cells. The targeting construct is then transfected into cultured ES cells.

ES cell lines are derived from the inner cell mass of a blastocyst-stage embryo. Homologous recombination occurs in a small number of the transfected cells, resulting in introduction of the mutation present in the targeting construct into the target gene. Once identified, mutant ES cell clones can be microinjected into a normal blastocyst in order to produce a chimeric mouse. Because many ES cell lines retain the ability to differentiate into every cell type present in the mouse, the chimera can have tissues, including the germ line, with contribution from both the normal blastocyst and the mutant ES cells. Breeding germ-line chimeras yields animals that are heterozygous for the mutation introduced into the ES cell, and they can be interbred to produce homozygous mutant mice.

As shown in Figure 1, the targeting construct used to replace the endogenous mouse C5aR sequence with the human C5aR sequence contains two regions of homology to the target gene located on either side of a positive selectable marker (PGK-Neo, a hybrid gene consisting of the phosphoglycerate kinase I promoter which drives the neomycin phosphotransferase gene). Homologous recombination proceeds by a double cross-over event that replaces the target-gene sequences with the replacement-construct sequences. Because no duplication of sequences occurs, the normal gene cannot be regenerated.

When the targeting construct is linearized, the neo gene is flanked by two regions of homology to the target gene. Selection of the cells using drugs (e.g., G418) eliminates the great majority of cells that have not stably incorporated the construct.

Although many gene inactivation approaches involving homologous recombination still use constructs that leave the positive selectable marker in the genomic DNA, it has become increasingly clear that this can cause a number of unanticipated effects. For example, the presence of the neo gene, often with its own promoter, can alter the expression of neighbouring loci. This can be a particular problem in gene clusters where neighbouring genes are in the same family, since the genes affected may have similar or identical functions. As a result, slight differences in targeting constructs have led to marked differences in phenotype.

For this reason, the targeting construct exemplified herein includes loxP sites flanking the PGK-neo gene, so that the selectable marker can be removed after targeting by transient expression of the Cre recombinase. This will leave the small loxP site in the

genomic DNA, but the construct can be engineered so that this is in an innocuous location (Figure 2). Although theoretically even a loxP site could cause alterations in the expression of neighboring genes, no such cases have yet been reported. The efficiency of Cre recombination from transient expression reported in the literature varies widely, from ~2% to ~15%. This rate should be distinguished from the efficiency of Cre recombination *in vivo*, where the expression of Cre is derived from sequences integrated into the genome and therefore will show longer-lasting expression in nearly all cases.

Figure 3 shows the mouse genomic DNA sequence (~22kb) encompassing the C5aR (C5r1) gene. Figure 4 shows the human C5aR cDNA sequence and Figure 5 shown the human C5aR protein sequence.

The mouse genomic region as shown in Figure 3 (the target locus) is characterised as follows:

exon 1: nucleotides 757-784 (5' untranslated region)
exon 2: nucleotides 1048-1152 (5' untranslated region plus start codon)
exon 3: nucleotides 11681-12733 (all coding sequence except ATG).

Figure 6 shows a restriction map of the 22kb sequence shown in Figure 3. The relevant restriction sites are as follows:

Enzyme	#Cuts	Positions:			
EcoRI	2	19417	20131		
EcoRV	4	199	5337	9874	19919
NdeI	2	8421	19713		
XbaI	4	1351	16923	17807	21857

In a preferred embodiment of the invention, the targeting vector used to generate the knock-in mice includes regions homologous to approximately 3kb genomic DNA either side of exon 3 (i.e. from about nucleotides 8000-16000 as shown in Figure 3). This means that following integration, the endogenous mouse exons 1 and 2 remain in the transgenic mammal but exon 3 of the mouse locus has been replaced with a sequence encoding human or humanised C5aR.

Example 2: Transfection and culturing of mouse embryonic stem cells

A basic protocol for the culture of embryonic stem (ES) cells and for introducing the targeting construct into ES cells is provided below. This protocol also outlines the method for identifying clones in which the target gene has been altered by homologous recombination. The resulting homologous recombinants are heterozygous for human C5aR and can be used to produce homozygous cell lines.

Materials

Targeting construct
Embryonic stem (ES) cells
ES/LIF medium
Trypsin/EDTA: 0.25% (w/v) trypsin/1 mM EDTA (20 mM HEPES, pH 7.3, optional)
ES medium
Electroporation buffer
G418
Freezing medium
Digestion buffer
Saturated NaCl
1% agarose gel

Transfect construct and select ES cells

1. Culture ES cells in ES/LIF medium. Passage cells every 2 to 3 days by seeding a 100-mm gelatin-coated tissue culture plate with 1-2 x 10^6 cells/plate. Leukemia inhibitory factor (LIF) prevents ES cells from differentiating. Passaging cells at a higher density may be preferable if blastocyst injection of the cells is planned (e.g., 1.5 x 10^6 cells per 25-cm^2 flask).

2. Harvest ~5 x 10^6 to 1 x 10^7 cells by adding trypsin/EDTA and incubating for ~5 min until cells are freed from the plate surface. Dissociate to single cells by pipetting up and down five to ten times. Add 5 ml ES medium. Pellet cells and resuspend the cell pellet in 1 ml electroporation buffer in the same tube. Typically, 10^7 cells can be obtained from a near-confluent 100-mm tissue culture plate.

3. Add 1 pmol linearized, sterile construct DNA.

4. Electroporate the mixture at 450 V and 250 uF in a 4-mm electroporation cuvette. Incubate 10 min at room temperature. Many electroporation conditions can be used with ES cells.

5. Plate cells in ES medium at ~2 x 10^6 cells per 100-mm gelatin-coated tissue culture plate. Incubate 24 hr.

6. Begin selection by changing ES medium to ES/LIF medium and adding G418 to 0.2 mg/ml and GANC to 2 uM (final).

7. Continue incubation, changing medium daily using ES/LIF medium and adding G418 (0.2 mg/ml final), until single, isolated colonies are visible (typically 1 week after electroporation). Remove an individual colony from the plate using an autoclaved pipet tip, and place in a 35-ul drop of trypsin/EDTA for 5 min. Pipet up and down about five times to dissociate cells. Transfer cells to a well of a gelatin-coated 24-well microtiter plate containing 1 ml ES/LIF medium.

8. Incubate until colonies are visible, but the cells are not differentiating (typically 3 to 4 days). Passage half of the cells to a well of a clean gelatin-coated 24-well microtiter plate. Add the remaining cells to 0.5 ml freezing medium and place at -70°C. Freeze overnight, then transfer to liquid nitrogen. Undifferentiated cells grow in smooth, round colonies. Differentiated cells are flatter with distinct intercellular boundaries. Proceed immediately to step 9 after placing half the cells in the freezer.

Screen for homologous recombinants

9. Incubate ES cells in 24-well microtiter plate (step 14) to near confluence (usually 2 to 3 days). Because it is not critical to prevent differentiation of the ES cells at this stage, LIF can be omitted from the culture medium; however, the presence of LIF may help to maintain cell growth.

10. Add 300 ul digestion buffer to each well. Transfer well contents to a 1.5-ml microcentrifuge tube, and incubate overnight at 55°C.

11. Add 150 ul saturated NaCl and vortex vigorously (the solution will turn milky white). Add 2 vol of 95% ethanol (the solution will turn clear except for precipitated DNA). Some investigators precipitate the DNA using 2 vol ethanol (or 1 vol isopropanol) without adding salt. However, the DNA pellet resuspends more easily if salt is added.

12. Resuspend DNA pellet in 50 ul water. Determine DNA concentration by measuring the absorbance at 260 nm.

13. Digest 10 ug DNA (or 10 ul if DNA concentration was not determined) with the appropriate restriction enzyme.

14. Fractionate the digested DNA on a 1% agarose gel. Transfer to a nylon membrane, and hybridize by Southern blotting to a suitable target-gene hybridization probe to distinguish the unaltered target gene from a target gene that has undergone homologous recombination.

15. Select ES cell colonies that show two hybridizing fragments of approximately equal intensity—one fragment of the predicted size for the unaltered target gene and one fragment of the predicted size for a target gene that has undergone homologous recombination. If the two fragments are of unequal hybridization intensity, the cell population may not be clonal. Freeze cells and store in liquid nitrogen.

16. If desired remove selectable markers that are flanked by loxP sites by transient expression of Cre.

The standard method for generating chimeras with ES cells uses 3.5-day-old mouse embryos (blastocysts). Embryos at this stage have a large fluid-filled cavity into which ES cells can be placed by injection. Blastocysts have already moved out of the oviducts and are found in the uterine horns. For injections, they must be harvested prior to hatching (loss of zona pellucida) and attachment to the uterine wall. C57BL/6 mice are used most commonly to generate blastocysts for ES cell injections. This strain yields reasonable numbers of embryos. It tends to produce high-grade chimeras that can be easily distinguished by coat color when used with ES cells derived from a variety of 129 sub-strains.

Example 3: Mouse model for K/BxN serum-induced arthritis

K/BxN TCR transgenic (tg) mice express a transgene-encoded (T cell receptor) TCR reactive to a self-peptide derived from the ubiquitously expressed glycolytic enzyme, glucose-6-phosphate isomerase (GPI), presented by the MHC class II molecule Ag7. These animals spontaneously develop a very aggressive form of arthritis, beginning at 3 to 4 wk of age. As in humans, the disease is chronic, progressive, and symmetrical, and it exhibits most (although not all) of the clinical, histological, and immunological features of RA in humans. Histological features include leukocyte invasion, synovitis, pannus formation, cartilage, and bone destruction. The murine disorder, critically dependent on both T and B cells, is joint specific but is initiated, then perpetuated, by T, then B, cell autoreactivity to a ubiquitously expressed antigen, GPI.

Strikingly, transfer of serum (or affinity purified anti-GPI IgGs) from arthritic K/BxN mice into healthy animals provokes arthritis within days, even when the recipients are devoid of lymphocytes.

Serum Transfer Protocol and Arthritis Scoring.

Sera from arthritic K/BxN mice at 60 d of age are pooled and injected intraperitoneally, in 150-250 µl total volume (7.5 µl serum per g weight, adjusted with PBS if necessary) into (knock-in) mice at days 0 and 2. Arthritis is scored by clinical examination daily over a 2-3 week period. A clinical index is determined over time: 1 point for each affected paw; and 0.5 point for a paw with only mild swelling/redness or only a few digits affected. Ankle thickness is measured by a caliper, ankle thickening being defined as the difference in ankle thickness vis a vis the day 0 measure.

Histology

The basic procedure of fixation, decalcification, paraffin sections, and hematoxylin/eosin staining of joint sections are as described (Kouskoff et al., 1996, *supra*). For immunohistology, unfixed and undecalcified cryostat sections are obtained. In brief, dissected ankle joints without skin are embedded in OCT, frozen in dry ice isopentane, and mounted on a cryomicrotome support at -25°C. After trimming the tissue block to a desired level, transparent tape is fastened onto the section surface

of the block. Sagittal sections (6 or 8 mm thick) are cut underneath the tape, and the tissue is subsequently transferred to an adhesive-coated slide. Slides are stored at -80°C until use, then acetone-fixed for 30 s to 1 min and air dried for 30 min. Nuclei are counterstained with 50 ng DAPI (Molecular Probes).

Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is solely for the purpose of providing a context for the present invention. It is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present invention as it existed in Australia before the priority date of each claim of this application.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

Dated this 24th day of December 2003

G2 Therapies Ltd
Patent Attorneys for the Applicant:

F B RICE & CO

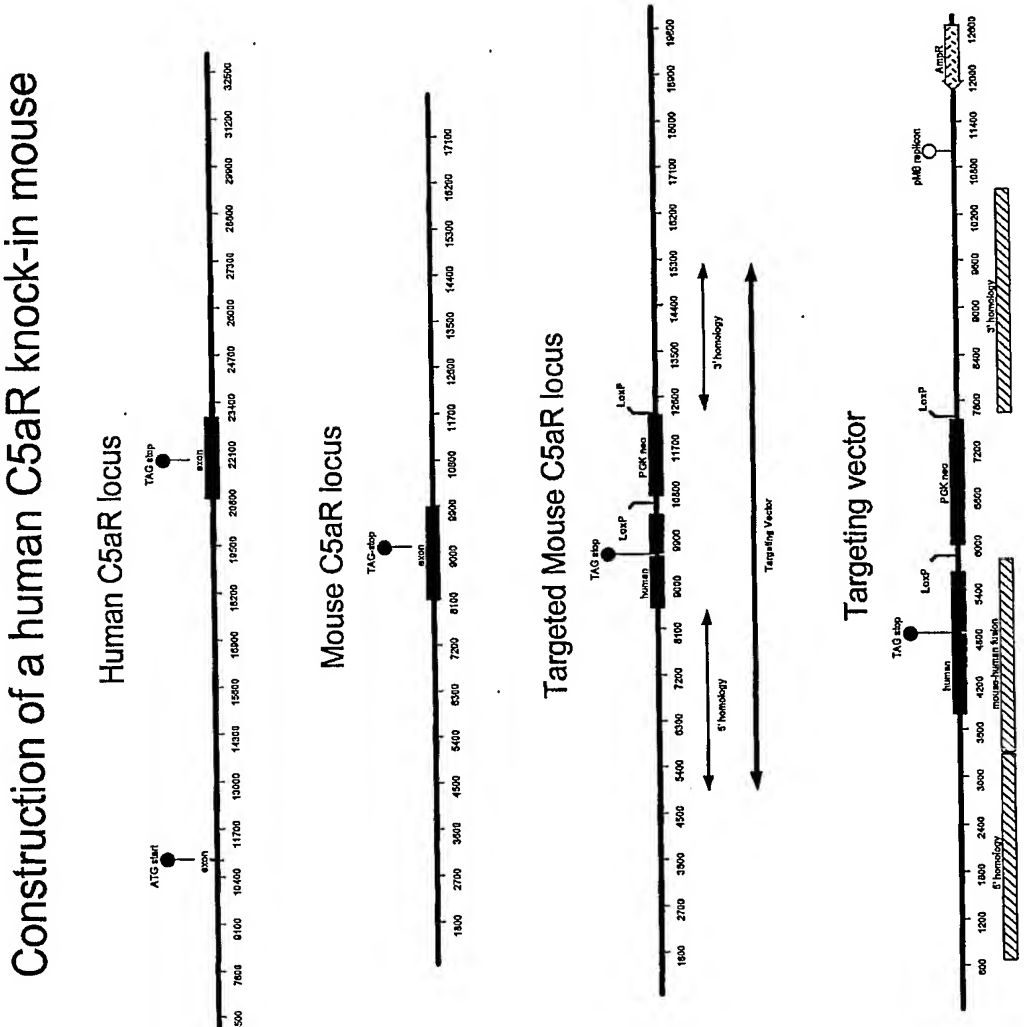
Construction of a human C5aR knock-in mouse
Human C5aR locus
ATG start
TAG stop
exon
Mouse C5aR locus
TAG stop
exon
Targeted Mouse C5aR locus
TAG stop
LoxP
PGK neo
LoxP
human
5' homology
3' homology
Targeting Vector
Targeting vector
TAG stop
LoxP
PGK neo
LoxP
AmpR
pMB replicon
human
mouse-human fusion
5' homology
3' homology

Figure 1

Final human/mouse C5aR gene locus after deletion of the PGKneo gene by Cre recombinase

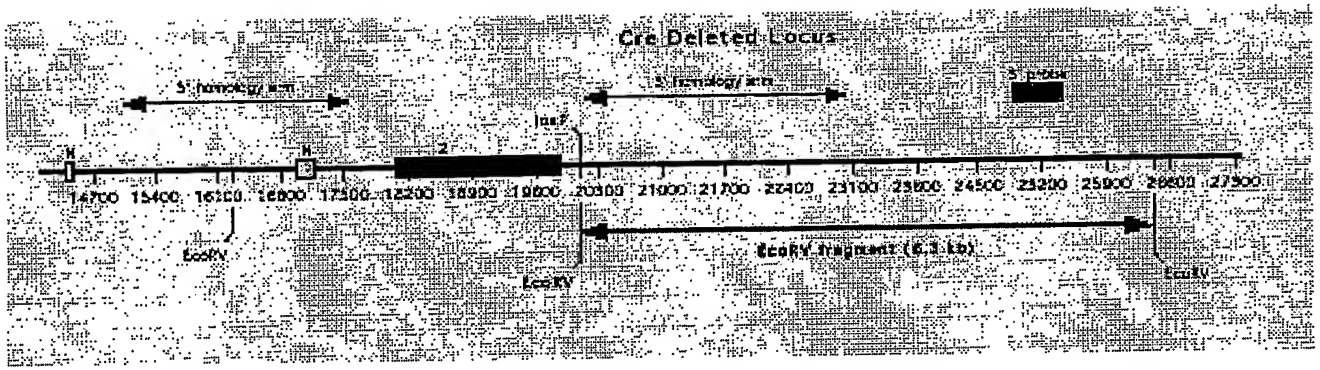

Figure 2

```
tttatttta ttttttaaa aaattggtcc ttcctatgca ggtggcctgg aatttttaat      60
cctcctgctt ttgtccaagt aatagaatta caggcatgta taattgtgcc taacctgagc     120
caatttgtc tttgctaaaa agcacaggtt ctcaacctgt gggtcctgac ccctgggggg     180
agggtgtcac atctcagata tcttgtatat ctgatattta cattacagtt atgaagtaac     240
aatgaaatgc ttttgtggtt gggggtcacc acaatatgtg gaaccaactg ttttcagggt     300
cacagtgtca ggaaggttga gagccactgc tacaaaggat ctcacaaagc ctactctgga     360
ttagaagtta ctgtgcagcc aggaggatgg ctttgaactt attctcctgc ctcagccgcc     420
tgggtgctgg ggtcggatgc agcactacat caggttttat gcggtgctgg gaatggtatc     480
cagggctttg cacatgctag gcaagtgttt aaccaaccaa gccatgatcc cagcatgctt     540
tgctttatta tttgagacag gtcctgttct gcagcccagg gtgatcctcc tgccccagtc     600
tcctgagtgc tagcattgag ttaacacatc tccctaaccc ccttaagaga aaacgccaag     660
accttggcca tctcttcagc cctctgtgtg ctttcttgct ataaaagcca ccaggctggg     720
gagctgtggt cattatttct gtcatgtaga aaccccagaa actccaaaac ttccttcaga     780
agaggtaggc tcctcctcag attggggaac gaggccagga aaagcagctg cgtccccaaa     840
agtgaagaag tcctggaaat tgccttttcc ccttctgggg ccagagactt ccttcctttt     900
ccaagttgac atctctcccc tggctggttg gtactgggtg gtgctgaggg tgtactgggt     960
aagcaccggt ggagggagcc tcagctagga tggtcagtga gtgaccaatg agcacctcca    1020
ggagacaaga cagtcatttc ctctcagttg cctgcatctc ttcttgaggg tttaaaaggc    1080
acagcctggg tgacagggac cttcaggcat ccgtcgctgg ttaccacaga acccaggagg    1140
agccaggaca tggtgagtgg attcctctcc ctgtctgact ttctttgccc atttctagct    1200
cctttcccat cctgagctca cactctgaga tgggatgtgg ccaacggact aaggggattt    1260
atgggaacca cggcggcctc accaagatgt aggctcaaga aggttcttca ggacaggaga    1320
ccctggagtc agctctcctc actgaagagt tctagaagtt gggcatgttc atttacactt    1380
gtaatctgag ccctcaagag gccggggcag gatgctgcag tcaaggctac atagtgactt    1440
ttaggctagt gtgagctaca ctatgagata ctgtgttcag agacaaatgg gctggaggta    1500
gagatcaggt gtgagagtgc tctgtgggag agctggaagc catgggtcca atgaccagca    1560
cctcgtacag ctgggtggga gtgattggaa attcaatctt acatagtaat tttgaggtta    1620
acctgggcta cactacatga gaccctactt cagaaaagca agaacaaaaa aataatttta    1680
caaactagcc aaggtggtgg ctcaaacctg ctatcctggc tcctagcagg aggtttgaga    1740
tttggggcga gcctaggcaa ctttgttgag actttgtctc aaagactaaa agcaaaacaa    1800
aacaaaaact caaacagtgt gaataaagga aggaagaatg aaacaattgc agaaacctgt    1860
tgggattgta gctcactgcc tagcctgagt gtggcccagg gttccgtctc ctacgctgag    1920
tctaaaacta ccaagcagag actgggtgct gtgacgcaca ccttttaatc ccgcactcag    1980
gaggcagaat cgggaggttc tctgtgagtt cgaggccagc ctggtaaaca tgtaaagaag    2040
tctaaggaag gtcaatgttg agagtcttga cagccagttt gaaagaacgc ccattcccag    2100
aaaagttaga ggaagcccag atgggagcac tgatggcctg ggtccttctg tggttaatgg    2160
ccatgaccct ctaggcaggt ccctctccat gcctgggacc tgacgttgag gcatggtgct    2220
agaccagcgg cgacttggcc ccactgtaac agaggatacg gtcttgcttc atccacacaa    2280
aagaggaaac ggaaaacttg atgacaggga gggtacacgc tttcttcatc cttcttctgt    2340
cccatccaat cctgtgtctg ccccgagcaa ttggggtttc cagaacaggg tgggttcttt    2400
ttcctttcta cacaacgttt ctgaagacga agtcacttta ttgaccaccc gaactgtaga    2460
gtccctgatt tgggctgggg cgtgactgag ttttttgttt tttgtttgtt tgtttgtttg    2520
ttttaaatac tggaataggc tgtcagtatc ttttttttct tttaagattt atttattata    2580
tataagtaca ctgtagctgt cttcagacac tccagaagat ggcatcagat cttgttacag    2640
atggttgtga gctaccatgt ggttgctggg atttgaactc cagaccttcg gaagagcagt    2700
cgggtgctct tacccactga gccaactcac cagccctttt tttttttttt tttttttttt    2760
tttttttaa agatttactt attttatata tgtgagtagt attgtctctt cagagacacc    2820
agaagagggc atcagaccca attacagatg gttattagcc accatgtggt tgctgggaat    2880
tgaactcggg acctttggaa gagcagtcag agctctaaac cgctgaacca tctctgcagc    2940
ccgtgactgg attcttaggc cagtagtcta tggctaagct atgcccctca cccctcactg    3000
ggggattcta ggcaggggct ctaccactga gccacactcc cagcccctca ctgggggatt    3060
ctaggcaggg gctctaccac tgagccacgc ccccagcccc tcactggggg attctaggca    3120
ggggctctac aacatttcag tccttgatct tttaagacag gatgtcacta tgtagcccaa    3180
tggtctaaat cacatgatta tcctcaggct ccctggtgct gggatcacag gcatatacca    3240
ccgtggctag cccctaaaca taatttttct tttgaatgaa taatttttt cttttggttt    3300
ttcaagatag gatttctctg tgtagccttg gctgccctgg aacttgctct ataaaccagc    3360
ctggcttcaa actcacagat cctcctgcct ctacctcctg agtgctggga ttaaaggcat    3420
gtgccatcac tgcctagctt tgaatgaata ctttttttta atattgtgaa taggcattta    3480
ctgagtgctt attgtatgct agtcctcttg ctaagcactt tagatttact acatagcaaa    3540
ctatcaataa aggagctgta gaatatccat gtatttcaag ggcaacacag cctttgaaca    3600
gacatatact atcccaatgg cattccacgc attaggcggt ataacctttt aaagagaagg    3660
ctcttgggat tcggccccac ccctgctctt gctgataggt tttgggaggc tttctaacta    3720
acctagagcc ccacttttta aaatctgtag agtgggtgtg gccatagtag cagcccaatg    3780
agggttgcat gtgttaaatg aagaaaagag cagttgaaag cccctcacaa gtggcccata    3840
cctgtaatcc cagcactcag gagaaagagg ccctgtctca aaagaaaata caaaaagcat    3900
gtaaacttat ggaccaggct aattatttta ttttgttttt ttaaaaaaga tttatttatt    3960
tatttattac atgtaagtac actgtagctg tcttcagaca ctccagaaga gggagtcaga    4020
```

Figure 3

```
tctccttacg aatggttgtg agccaccatg tggttgctgg gatttgaact aaggaccttc    4080
agaagagcag tcaggtgttc ttacacgctg agccatcgca ccagccccag gctaattatt    4140
gttattttga aataggttct catgtagata aggctgaacc tagaactcac tatgtagcca    4200
aggatagctt tgacttcctg tcctcctgct ccacctctgg tctctctctc tcgatatata    4260
cacacatata tgttcatttt atatattata ttgtataata gtttatagtc ttctttttt     4320
ctttttttt ttttttttg ggttttcaag acagggtttc tctgtgtagc cctggctgtc     4380
ctggaactca ctctgtagac caggctggcc tcgaactcag aaatccgcct gcctctgcct    4440
cccgagtgct gggattaaag gcgtgtgcca ccacgcccgg ctatagttca tattctttaa    4500
gcaactattt ttatatcatt tatttatttg tcttacaaga tttgttttta attgtgtgta    4560
cgcttttgag tctgtctgtc acacgcgtgc agatgccctc agaggacaga aggtgttgga    4620
ttttcggagc tggagtttca ggcagttgtg agatcccctt gggtgctgag aagtgaacac    4680
atgtcctctg cggaagctga cagtgctctt cattgctgaa ccatctctcc acttccttct    4740
tagtcttttt tttctcaatt gtttttctcc ttaaaaaata ttttgacctt atgattagtt    4800
gagtccacag acatggacac tgtgtatacg gagggacaat gacatcttct ataatagttc    4860
aaattatgta tgcatataat atgttacata tattgtattc cacatccaag aaccatataa    4920
acaggagaaa gtgctctctc tctctctctc tctctctctt aaagatttat ttatttgtta    4980
tatgtaagta cactgtagct gtcttcagtc actccagaag agggcatcag atctcattat    5040
ggatggttgt gagccaccaa gtggttgctg ggatttgaac tcaggacttt ccaagagcgg    5100
tcagtgctct tacctgctga gccattctcc agcccaggag aaaactctct taattcccga    5160
gtcccagtcc cttccttaga ggcagccact actgtcagta tgtgaggcta gtctgtatgt    5220
acacgtgaat ggacacacac acattcatgc acatgttgtg catccttcct tctacaggta    5280
atatacattg tgcattgtcc actgctattt cctattttaa acaaagtttc cctagatatc    5340
aaagaccttt ctgtatcatt tcattaacaa ctaccttatt cttttaaaaa ctgcatagtg    5400
ctttgttgcc tagaggctat tactaagtta ctcagtccct tttggtcagt tcaaaatatc    5460
attatcatct cctcctctgc ctcnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    5520
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    5580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    5640
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    5700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    5760
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    5820
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    5880
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    5940
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    6000
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    6060
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnntg tatctggaag    6120
ccccaggaaa ccacagagtt tacgatcagc atctttttct ctgcctcatg aagtcacgaa    6180
aaggacaggg tgagatcatg tcaggaagca agaaaggaga aggtcagccg gcaaggatga    6240
gagatggggt tagagaggcc cgtgtcagaa gtctgagtca ttgtagtaag gatggtgaga    6300
tctgcagatg ccaggagaag cattccaccc tgtctggggc cttgagaata cccagagggc    6360
aggctgtgag gtttcctata gggcccagaa ttaatcctca agtgacctga gcatgggacc    6420
ctggggatgt ggggatgccc agagtaacaa gtagaaagat acagaactga gggtgagcca    6480
gagtgaaatg agtggctggg tcctgggtct gtctgtctgt ctgtctgtct gtctctctct    6540
agctttcttt ttgtttttct ctgtgtagcc ctggctgtcc tagaactcac tctataaacc    6600
cggctgacct caaactcaga gatctgtcct cccctgccac ccgtgctggg attgaaggtg    6660
tgcatcacat caccaccctc ctgcttgaaa tatttttaaa ttatagaaaa gtgtgaggct    6720
agtacaagaa gtttatttat cttcttttgt tgtcgttgtt atttattatt actgagatgg    6780
ggtctcgcta tgtagctcag gctgacactc aatgtaatgc atagcccagg ctggtctggg    6840
gcgccacatc ttcctgttca gcctcctcag tgttggcatt acaggcgagc attaccataa    6900
attcttgtgt ttcctcccca agagtcccca catgcagtct atggtgtata tatgtttacc    6960
catgtatata cttataaatc tttgtatgta tctatgtttc tgtttctata aacatatcag    7020
tttagctgca tatcgacata tctgtttttc tatctgtcaa tatagctatc aattatgctc    7080
catccattgt ttctctacca ttcatctcta gctattcatc atctacccag gcattttcta    7140
ttcttctatt agtttagcct ggtcttgaac ccccaaccta gctgaggata gctcttttt     7200
ttttttttat aatggtactg aagtttattt tttttaaaga tttatttatt tattacatgt    7260
aagtacactg tagctgtctt cagacacacc agaagagggc gtcagatctc gttacagatg    7320
gttgtgagcc accatgtggt tgctgggatt tgaactctgg accttcggaa gagcagtcgg    7380
gtgctcttac ccactgagcc atctcaccag ccctgaggat agctcttaac tcttgatctt    7440
cttgcctcca ccacatagat tcaggggtta caggtgcact accgtgtcca gtctatgtag    7500
cactggggat ggaacccaag ggttcataca tgataaatga gcactctaca agatgagcta    7560
tgttcctaac ccatctgcct gtctttctat cacctatgtc ccatccgttg atctataaat    7620
ctttttattc atctttcacc cacccaccca accattcacc cattcattct tagtaaccaa    7680
ggctggtctt gaactcttga ttttcccacc cgagtcttcc cagtcctgag atgatacaag    7740
tgcacatcac catacccgtg tcaaaatcta cttctaattc ttatttctgt ttttaaaaag    7800
aaaagttatc tgttttatgt atatatgtgt ctgttgagtg tatatatgtg caccaagtgc    7860
ctacaggagc ctgcaaagat cagttgtcag attctgccat tggagttcta aacagttgcg    7920
agctgtcaca actctggtcc tctacaggag cagcaactgc tcttaactgg tgacccatct    7980
```

Figure 3 cont...

```
cttctaattt ctctctctct ctctctctct ctctcnnnnn nnnnnnnnnn nnnnnnnnnn  8040
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn  8100
nnnnnnnnnn nnnnncctct ctctctctcc ctcgctcgct cgctctcttc cttcgttcgt  8160
cgcttcgttc cgcgcgtcgt tcgttcgttc cttcattcgt tcgttttgtt ttttgttgtt  8220
ttgttttttt ttaattaggt attttcgtca tttacatttt caatgctatc ccaaaggtcc  8280
cccataccca cgcgcgcaat cgcgtaacca cgcactcgcc ctttttggcc ctggcgttcc  8340
cctgtactgg ggcatataaa gtttgcaagt ccaatgggcc tcttttgcca gtgatggccg  8400
actaggccat cttttgatac atatgcagct agagacaaga gctccagggt actggttagt  8460
tcatattgtt gttccaccta tagggttgca gttcgcttta gctccttggg taatttctct  8520
agctcctcca ttgggggccg tgtgatccat ccaatagctg actgtgatca tccacttctg  8580
tgttttgttt gttttttga gacagggttt catcgtgtag ccctggctgt cctggaactc  8640
actctggaga ccaggctggc ctcgaacaca cagggatcta cctacctctg cctcccaagt  8700
gctgggatta aaggcatgct ctaccaccac ctggctagtt cttatttctt attttgcctt  8760
ttgctggccc caaatacttt gcctccactt ccaattgtaa gtcccaaaac ttagggtttg  8820
gaaaatgggt ggcttgctag actgtcaagg agataatgaa ggaagaaagg gaggctcaga  8880
gccagagaaa tttgcaaagg aacctgtatg ccccataggt ctggatcaca ggggataact  8940
ccaaagccag tatccaaagg acagcctatc ctcagctggg ggtggagtct ttgcctgctt  9000
cccgcctata ctaaaatgtc gaactatttt ttctctttct ctctctctct ttctattttt  9060
tcttttaaag atttatttat ttattttatg tatatgagta cactagggca ttagatctca  9120
ttacagatgg ttgtgagcca ccatgtgatt gctggaattt aaactcagga cctctaggag  9180
agcatccagt gctcttaacc tctgagccat ctctcagccc tctggttttt tgttttgttt  9240
tgttttgttt ttttgttttt ttaaagacag tatctcattg tctttgttag cagctctgtc  9300
ctgtcctaga actcactatg cccaacacac tgacctcaaa tacatgctta tccacctggc  9360
tctgcctccc aaatgctggg attaaacatg tgtaccacca ctacctggca tctctgtcca  9420
tcattttaat caaagagaaa aatgtataaa acttttttctt aagtagccca gactggtaat  9480
gaggatcact gtgcatctga ggatgagtct gttgcctctg tctcccagat tctggggtgg  9540
gagtcaccat ttctagttta atgttgtgct gggtttggag tctgtggctt catgctttga  9600
gactggtttc ctgtaaggca ggcgatcatt gaatgcttcc ccaccctgcc tcctgttact  9660
gaatgtcagg attgtagcca tgagccgcca tgcccgcttt aatataagat catttaaagc  9720
agtagttctc aatctgttgt cagagagtag caagcttaca gtcaggaagt agcaacaaaa  9780
gtcattttat ggctgggggt caccacaaca tgaggaactg tgttaaaggt ctcagccttc  9840
ggaaggttga gaaggaaacc tcaaacccac agatatcggt cacagttctc aaaggaccac  9900
attgcccaat atgtttatac accatggtca catttccagc ccaccgagga caccaggata  9960
aagcttcact gccaacaatg aggtgtttca aaattagatg tcattgtcct gtctttatac 10020
caactttggg ttttagtcca aattcagggc atacacatct ataattctag cacacaggag 10080
gtagaggcag ggggatcagt agtttatcat cttgagctac atagtgagtt ttgggactag 10140
cctggtatac agtggattct gtcaaaaaac taaatgacaa agaagtaaca acaacaacaa 10200
caaaataata ataataataa taataataat aataataata ataatattat tattattatt 10260
attattatta ttattattat tattattatt tgtgtgtgtg tgtagtgtct ggacacatag 10320
gtcaagctga gcttgaactc aggacaatcc tcatacactg ttttgagctc tttatatcac 10380
tgggagctgg agagtgtagc tcaggagctc aacagtacct gccagagtga caggagttca 10440
gttccaagca cctatgtaga gtatgctcac aaccagatgt aattccaaag tgttcaatgc 10500
cctcttctag cctcccaggg cacgctcctc tctctctctc tctctctcct ctctctctct 10560
ctctctctct ctctctctct ctcnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 10800
nnnnnnnnnn nnnnnncccc ccatacagta ccaatggtaa gatggtttag caggtaatcg 10860
cccaagcctg gagacctgag ttctatctta ggacccacat aaaggttaag ggagagaacg 10920
cagtgcacaa agttatcccc tggctttcac atgtgtgtta tggcatgcac atgcatacat 10980
acatacatgc atacatacat acatacatac atacatacat acatagacag tgacaaatta 11040
aaataatacc tcattggtca gtcactgcac ccctttaatc ccagcactca gaagccagag 11100
gcagttggaa ctctgtaaga gtggagccag cctggtctac agagtgagac ttttttcttt 11160
tttctttttt ttttaaagat ttattcattt attatacgta agtacactgt agctgtcttc 11220
agacactcca gaagagggag tcagatctcg ttacggatgg ttgtgagcca ccatgtggtt 11280
gctgggattt aaactcctga ccttcggaag agcagtcggt tgctcttacc cactgagcca 11340
tctcaccagc cccgagattt tttctcatca cctccctacc ccaatccata tacttgatta 11400
aagcccaggt ctggagagcc atgcctgtag tctcagcatt gggcagctga agtagatgga 11460
ccaccatgat ttcagtttat cctgggcttc agagtgagtt taagaccagt ctgggtaatt 11520
taacagagaa cctgtctcaa aataaaatct acaaactata ctagttttat aggtgttcag 11580
catcccttgg tagagttgag actcagaaag acgggcaatg cctccatccc ctgggaatgt 11640
gtctaccaac tcacacaatc tacctgtttg atttgcttag gaccccatag ataacagcag 11700
ctttgaaatc aactatgatc actatggaac catggatcct aacatacctg cggatggcat 11760
tcacctcccg aagcggcaac ctggggatgt tgcagccctt atcatctact cggtggtgtt 11820
cctggtggga gtacctggga atgccctggt ggtgtgggtg acagccttcg aggccagacg 11880
ggccgtcaac gccatctggt ttctgaatct ggcggtggcc gacctcctct cgtgcttggc 11940
```

Figure 3 cont...

```
actgcctgtc ctgttcacga ccgttttaaa tcataactac tggtactttg atgccaccgc   12000
ctgtatagtc ctgccctcgc tcatcctgct caacatgtac gccagtatcc tgctgctggc   12060
taccattagt gccgaccgtt tcctgctggt gttcaagccc atctggtgtc agaaggtccg   12120
cgggactggc ctggcatgga tggcctgtgg agtggcctgg gtcttagcat tgctcctcac   12180
cattccatcc ttcgtgtacc gggaggcata taaggacttc tactcagagc acactgtatg   12240
tggtattaac tatggtgggg gtagcttccc caaagagaag gctgtggcca tcctgcggct   12300
gatggtgggt tttgtgttgc ctctgctcac tctaaacatc tgctacacct tcctcctgct   12360
ccggacctgg agtcgcaagg ccacgcgctc caccaagacg ctcaaagtgg tgatggctgt   12420
ggtcatctgt ttctttatct tctggctgcc ctatcaggtg accggggtga tgatagcgtg   12480
gctgcccccg tcctcgccca ccttgaagag ggtggagaag ctgaactccc tgtgcgtgtc   12540
cctggcctac atcaactgct gtgttaaccc tatcatctac gtcatggctg gccagggttt   12600
ccatggacga ctcctaaggt ctctccccag catcatacga aacgctctct ctgaggattc   12660
agtgggcagg gatagcaaga ctttcactcc gtccacgacg gacacctcaa cccggaagag   12720
tcaggcggtg tagaggagaa gccacaactg gcctagctgc tccttttcca gccctcctac   12780
cccctcctct tcttcctcct cctgcctctc ctccttcctt ccttccttct ctttgcatgt   12840
ttaattttct gcaattctct aagttgctct gactagcctt gagcccagga tcctcatgaa   12900
ggctgagatt ataaatataa attcctttga tgaaaagcat cacattaaga tagtactcgg   12960
ctttttttct aaggcttttt ttttttttct tggctacgtt gcccacctgc agtggctagg   13020
cagatacacc taatgatgac ctccaggggt tggataacag agaacaagag aatttcctgg   13080
ccttcttctt cctctcttcc tcttcttttt ccctcctcct ccttcttctc ctcctcctcc   13140
cttttttttt ttatggttct ggtctgaacc caggtctcaa tggaacccag ggcttatgga   13200
tatatcacat aagcaagcta cagccccaaa ccccaggcaa ccagtatcca cccacccttt   13260
atttcttttc tatgtttgat tttttttttt tttgagacaa ggtctcatgt agggtagtct   13320
ggccttgaac tccagatcct cctgtgacca tctcccaagt gtcgtgactg tagacctgtg   13380
ctgttgtgtc cgacctatcc tttatttcta caattttgtg ttttcaggaa tggtatttaa   13440
tggaacccaa catatccaag ctttgtaaaa acaactatgc atggcttact tgataaattt   13500
ttttttttta aaaaggtaca gaaatgtgtt gtttaacttt ttaaaagcac gtatttattt   13560
attttgtggg gggtgagggg gtggtgctgg gcaaatgtca tggtatatgt gtggaggtca   13620
gaggacaacc tgttgaaatt ggttctctca ttgcaaccat gaaggtcctc atggaatcga   13680
acccaggtca tcatacttgg cagcaaacac ctttacctgc tgagtcacat cactggccag   13740
aggtgtcctg tcttataatg cgttctttca gcttaatgaa tgtgtgtgca tgagtgtatg   13800
tgttggctag aaaatatgta cagatcaaca ccagaagtat catgcaagca tgggaatggt   13860
tttgaatttc ctggtcaaat taaaaatgtg aaagaagacc tgggtgtggt ggcgcaaacc   13920
tatatcccag catgtgggag gttcaggggc cagaattgag ttttagaacc cagcctggct   13980
tacccaggga gactgtctca tgagatccaa ataaacagta tatgatggaa aacactggag   14040
tttagctctg ctaggccctc tcttcttccc agtggatatg tgaccactgg ttgtcacata   14100
tcacagaccc agcctacctg tgttctgcta ttcacacttt ctatatgatg acactaacct   14160
cactgaattt ttacaggctc catgccttgg catttattat ttatgtattt atttattttg   14220
agacaggatc tctttacata gccctagctg tcctggaact cactatgtga accaggctgg   14280
cctagaactc acagagatgg gcctgcctcc atctcctgag tgctaggatt aaagacatga   14340
gccaccacat ccagctttat tctatgtttt gtatggcctc tatgagtttg aaacatttaa   14400
tcaattagtt agttaattaa ttaatatatg agatgggatc tcatgtagcc caggctagcc   14460
ttaagctggt tttacagctg aggtgggatt ataggtagtc ctcctgactc ccagttgtct   14520
ccctcttgtg gcttttctca ttatcggtca catctgtatt gccacagctg agcttctcac   14580
ccactgaccc atgccccagc tgtcccaaga acctcttcct ccccttgctt ttccattcca   14640
ggaaaaacca cactggcaac ctgctcaccc aggccctttc agctgcccca tcacagaccc   14700
agccctccct tcttaccaca cacccggctc tacatcctgc ccccccccgc cgcacccccc   14760
cccgcctcct tcatgcctct ccctttcctt gatctcctgg ttgcccagca cctcttccaa   14820
ggaccatcct gctcccatcc tgtcttcttg ccaggtgtcc cctccttaag ggagtccccc   14880
gtgacagccc tcagtttccc ataagcaccc taccatcaat ctttttctct ggctgcgatt   14940
gagcttcctg gttcagggag taagtagtag gtagggattc acctccttct ggccttgctg   15000
taatgagatg ctgttttaag ggttgggctg agggctgggg ctaggggtg gggtggggtt   15060
agaaagacgg atcagtgatt aagagcattt gatgttcttt tagagcagcg gttctcaacc   15120
tctgggtctc aacctctttg gcaaacttct gtttccaaat tatttacatt ccgattcata   15180
actagcaaaa ttacagtttt gaagtagaaa tgaaaataac tttatggttt gggggacac    15240
tgcagagtga ggaactgtat ttaagggtca taggtcgtag catcatgaag gttgagaact   15300
actgttttaa aggattagtt cagttcccag catccacata gtgtctccta atgatttgta   15360
atggctgcct tggacaccaa gcccacacat gctggacata catgcaagca aaacacccat   15420
acatataaaa ttatatataa tatgtaagct gggcccagga tacagtgttt cagttcagta   15480
ggtagcatgc tggcctaaca cgcacaagcc tctggttcag tcccctgcac tgaataaaat   15540
ctccaatagt ggttgggtgt ggtggtacat gcatttaatt ccagcactcc agacacagat   15600
gcaggcagac ctctgggagt ttgaggccag ctacttagtg agctccaggt cagtccaagt   15660
gagacttggt ttcaaaataa aacaaatata tacacacaaa gaaactaaat ctgcatggtg   15720
gatttaggag gtagaggcag gaggctcatc tagtcaagga gagtttgtgg ctagcctggg   15780
ctacatgagg ccattctggg ctacatgagc ctctgtctga aaacacaaac aaaaacaaat   15840
gaacaaacaa acaaacaaca acaaaaatcc cagccaggct tggtaacaag catttgggag   15900
```

Figure 3 cont...

```
acggccatag gtggacctcc gtgagttcag gctgcagaga gaggccagtt taaaaccaaa   15960
acgagacaaa aagggatgct cagtggttta agagcattgg ctgctgctcc tccaggggac   16020
tgaggtttcc ttcccagaac ccacagggca gctcacaact gtctgtagct ccagttccag   16080
gggagctgat gcagtttcct tgcctccaca ggcatggtgt gtagcacgca gatatacaga   16140
caaaccactc atgcaccaaa ggcaaaaata aattaatcta aaagaaagga aggaaggaag   16200
gaaggaagga aggaaggaag gaagaaagaa agaaagaaag aaagaaagaa agaaagaaag   16260
aaagaaagag aaagaaagac aggaaggaag gaaggaagga aggaaggaag gaaggaagga   16320
aggaaggaag gaaggaattg gacatacagc aggtggtggt catgttgaga gacccccacc   16380
ccaggtgact cccaggcagg tcagggttaa gcaacgcagc tcaaaacaga agtttgcaga   16440
gtccagggga ttgccaaatg tgtggcctgt ggaatctgct tatgtcaaca gggttggaag   16500
gggaagtgag caggaaagga agtgggctga gagcttggcg gactctagtg tgttctttct   16560
cctcccccag ccccagcctt ctggaccctt gggtcttaca cacctatctg ttcttcagat   16620
gcagggctcc aaggcctggg gccagagccg ccttcccttg taacggtgac ctccgggagc   16680
tcacatccag gaagctgtta cattgcagta gagtcttctg ggatgaaata tgaggggctg   16740
ggagacgggt cagtgagtaa agtgtttgcc atttaaacat aaggatatgc gttccagccc   16800
aggctatgga tttgcctggt acagaggcac ggtgggttgt gtttgtaacc tcagcacggg   16860
agagtgagac agatggatct ctagggcttg ctgaccagca ggcctgggtt aatcagtgag   16920
catctagagc aagttgagag ccttggtctc taaacacaag gtggaaggaa agggagggcc   16980
ctggagaggt ggttcatagg taccgctctc agcagcaagc actctcacct gaggagccct   17040
agccctagct ctactactga gccacactcc cagcccctca ttggtgagtt cttggttctg   17100
ttgagccagg cccccaatcc tttgctggag gattctaggc aaatatccta acactgagct   17160
gtgcactgct ccagaccttt tatcatcttg gcacatctgt tgaccaggta agtctcccat   17220
gttgaggtgt ggagaacact gaggcctttc aggatgagag agagagagga gaggcctgca   17280
tcacagaatc tgtagtgcct tgacccagaa gcaatttcct ctaacaacat gactttatgc   17340
tctaaatatc aacagaagaa tttgtgaccg catccttctc agccttaagc aaggctcaga   17400
gagaaagacg accatcagga actgctgagt gacgagagtc catgtcaggg ttgaggccat   17460
gtcctgctcg gtgtcctaag cctgcaccat gctgtaggtg tatagtttaa gacagtgtac   17520
tctagggcac actttaattc ccacaattgg gaagctgagg aaagcaaatc tgtgagtttg   17580
aagtcactct ggcctacgtg agaccctgtc tcaaacccaa cccaacccaa atcaaaccaa   17640
accaaaccca gccactatag ccaacttctt ttttgttctt gtcattacta ctactactac   17700
aaataataat aataaatatc taataataat tttcacttta aatatctgtg cacatgggcc   17760
tgtgagagtc acagtttgta tttgaaggtc agaggctagc cttaattcta gagctttcct   17820
ttctactttg agacagagtc tcttgttgct tgtaatggca aaggccagct ggcccacgtt   17880
tccagggatt ttgactccct ggctgtcttc ctttctgaac gctggcatca catacatata   17940
ctactgaatg tggctattat atgggttcca gaacttcaac ctcaggtccc catgcttgtg   18000
tgacgagcac attccccacc aatccaccca tggggaccgg acagatctct cccgtgagct   18060
ccccttgcc tctgcctccg gagtgctgga gtgacaagca tgttctccta tgcctgctgt   18120
cttcccattt tacaggtaaa aaaaccagag gcccagaaag gggacaggat ttgcttattt   18180
tggggcatgt gggggtttga gacagggttt ctctatgtag tcctggctgt ctctctgtga   18240
ctcttggctg gcctcgaact cagagacctg tctgagtgct gtgatcaaag gtgtgcgcca   18300
ccactgcatg acaggacttg cattttatgt tccccggaaa cctcaggccc tgggctcagc   18360
ttcttgatct ttctgaggag ggttcattct gggctatcat cctcacaaca tttgaggaag   18420
gaaagatctt taagagtctg tggctggcag gaatgagagg cagagaacag cgcagccggt   18480
cagtggaggg ttagcaggcc gctggtgatt actgcagaat cttaggggtc ctttagtgcc   18540
aagggtgggt gggaagtggt ttcagagata gccctccaga ccttgctgtt caaagcccac   18600
acacctctgg cttccaggaa gctgatagta gtgaggctgc gggtggaggc acacactttc   18660
ggcttttccg acctttctgt ctgtgggtta atttgtgact cacggggagg aagaaaagac   18720
aactatttcc ctggggctag cggaggccac gcctgttttt cctggttaag aaggttgcgc   18780
agggtcctca gagaatccca taggatctgg ggaagggttg cattgctgag actcaggccc   18840
gctactgtcc ctggtggaga gactctgggc ttccttgcgg ctgctgaggt ctgctgtgct   18900
tgtgcattcg gccaatttgg gaccagtcag aagagaggtg aggaagggag gcataaagga   18960
ggttgcgaga aggggtggag aggctcataa tgtttgcctt agaagctttc attttgaaat   19020
cttgggagtc agaattagca ttccagatta tatatgttgt attttcctga gacaagagct   19080
catgctgtcc aggctgacct caaactcact atgtagttga ggtgatctcg aacacccgag   19140
tgtcctgcct ccacctccag agtacaggga tcatcaaaca caggttatgt agtgctggga   19200
gcagagctca gggactttgg cactctacca actgagccac acccccagcc ctgaatcata   19260
taaaataatc tgtttcatta cgacatttat attatatatg aatgttcttg agttttgctc   19320
aaattcacca ccatctcttt tctcatcagc ttgtatgttg ttgttgttgt tgttattatt   19380
gatacaaaat atctctacgt agctctgact gtcttggaat tcgttatgta gacaaggctg   19440
gattcacaga aatccacctc cctctgcttc cagagcactg ggattaaagg catactcctg   19500
gcttatactt aaaagtggca atttggagct gaagagatgg ctcaatggtt aagaacatgc   19560
aatgttcttt cagaggtcct gagttaggtt ctcagaaccc atcttatcag tggcttacga   19620
acacctgtaa ctcctgctct agggagtcag atgccttctt ctggccccag caggtaactg   19680
cacacatgtg gccaacactt gtgtgcatag acatatgtaa aataatggca ataatatttt   19740
atgtattgtg tatagagcca aacaaatata aatgatttac tgtaaaagaa agcaatgtca   19800
ctgggtgcag aagtgtccat ttgtaatccc agatgagaag gcagataaga aaagaacggt   19860
```

Figure 3 cont...

```
ttcttttact ctctggccca tctgttgagc cagttggcaa acttgaggtt ctgtgagata   19920
tccagtctga aaaaatgtgg agggctggag ggtgtggctc agtggtagac cccctgccta   19980
gaatccccca gtgaggggct gggggcgtgg ctcacagccg gagccccttg ttaagctgga   20040
aagcgggaga tagcgcgaga tagagagggg ggtagacgga gagagagaga gcgagagaga   20100
gagagagaga gagagagaga gagagaacat gaattctggg aaccatcctt gtctctcttt   20160
acagaggaaa taccataggc tgatagtgac tgagtacaga aactgtccca gactattatc   20220
agtagtagct gtgaagggtg gggtcagaga tgggaagaga ggtagtgata acagcagttc   20280
acacacacat acacacacac acacacacac acacacacac acacacacac acacacaaac   20340
acacacacga gcaggcacac cctgtctgct gtttgctgtg gacgagcact gtggcagcct   20400
gtctccatag cagatccgct aactacactg actatccgca gcgctcgctc tcccaggtgg   20460
ggtctgtgat tatccctata caggtacaca gagattccgc agcttgttga aggccacaca   20520
gctattgaag ctttgagttt ttgtactctt gttatgctct atattgcttg ttttgtttgt   20580
ttgttttgag atgagggtct taacttatag cccaggttgg cctcaaaatc atggcatttc   20640
tcctgcttca gcctctgagt gctggggtga caggtgagtt tttgttttgt tttgttttta   20700
aacaggtaca gtttacttta aggaaggaaa actactcaga aaaatggctt ggcctcatag   20760
ctggctacct ggcagagctg agagtgtccc aatttccgtt ctgtccttcc tgttttaaca   20820
gtgttggcca aggctttggg cagtgccaga caacccataa atagtcagat gagagctgca   20880
ggttccagcc actccagaca tggggttggg tgtcccctcc cgcccaggtc ctgtccttcc   20940
ccgcctgctt tgtgtcttgt gtgtgtttct taggctttag ttcttctgtc ccaccaaact   21000
ggtgagctgg gtcctagagg aggatgtgca cagacagagc cagccgtgac tgcgggtcag   21060
ctcagggcca cggggataca cggctgacta gcttcccagt ttctcacatc tggggccggt   21120
aatatttctg gactccctag ggacacgctg caattcagtt ctgtcttctt agctgagtga   21180
ttttaactaa gttactcacc ctctctctgc ctctttagct gcagaatcgg cttaccaaga   21240
ctgtatcaaa acacagtgtt gaaaggtgtt tggggccagg ccttgcacgt gcacaaaatg   21300
gtgccctcta atatcctaaa actattatta ttatttttatt aggtaatttt gtgtcttagt   21360
tagggttttt attgctgtga agagacacca tgaggggctg ggggcgtgac tcagtggtag   21420
aacacccacc taaaattccc caggggcaca cgcaactctc tctctctctc tctctctctc   21480
tctctctctc tctgacaggg tttctctgtg tagccctggc tgtcctggaa cttgctttat   21540
agaccaggct ggcctcaaac tcacagagat ccctctgcct ctgctccaag tgctggaatt   21600
aagttgtaca ccaccactgc ctggctaatt atttctatct taatagtttc ttttcctgtt   21660
gcttgtgatg aaatactcca cagccagatg tggtggcaca ctttttttatc ccaagacact   21720
tgggaggcag aggtaggtaa atttctgtga gtttggggcc attctggtct acataaaata   21780
ctcttaaagg gctacttaag ggagaaggta cttattatag attattatat attatgtcat   21840
tatatattat atataatcta gagaattaat attataatat ttctataata catactatgt   21900
aatataatat taatatcaat acaattatat tatctactat tcattataca ttaatatata   21960
```

Figure 3 cont...

```
   1 aggggagcc caggagacca gaacatgaac tccttcaatt ataccacccc tgattatggg
  61 cactatgatg acaaggatac cctggacctc aacaccccctg tggataaaac ttctaacacg
 121 ctgcgtgttc cagacatcct ggccttggtc atctttgcag tcgtcttcct ggtgggagtg
 181 ctgggcaatg ccctggtggt ctgggtgacg gcattcgagg ccaagcggac catcaatgcc
 241 atctggttcc tcaacttggc ggtagccgac ttcctctcct gcctggcgct gcccatcttg
 301 ttcacgtcca ttgtacagca tcaccactgg ccctttggcg gggccgcctg cagcatcctg
 361 ccctccctca tcctgctcaa catgtacgcc agcatcctgc tcctggccac catcagcgcc
 421 gaccgctttc tgctggtgtt taaacccatc tggtgccaga acttccgagg ggccggcttg
 481 gcctggatcg cctgtgccgt ggcttggggt ttagccctgc tgctgaccat accctccttc
 541 ctgtaccggg tggtccggga ggagtacttt ccaccaaagg tgttgtgtgg cgtggactac
 601 agccacgaca aacggcggga gcgagccgtg gccatcgtcc ggctggtcct gggcttcctg
 661 tggcctctac tcacgctcac gatttgttac actttcatcc tgctccggac gtggagccgc
 721 agggccacgc ggtccaccaa gacactcaag gtggtggtgg cagtggtggc cagtttcttt
 781 atcttctggt tgccctacca ggtgacgggg ataatgatgt ccttcctgga gccatcgtca
 841 cccaccttcc tgctgctgaa taagctggac tccctgtgtg tctcctttgc ctacatcaac
 901 tgctgcatca accccatcat ctacgtggtg gccggccagg gcttccaggg ccgactgcgg
 961 aaatccctcc ccagcctcct ccggaacgtg ttgactgaag agtccgtggt tagggagagc
1021 aagtcattca cgcgctccac agtggacact atggcccaga agacccaggc agtgtaggcg
1081 acagcctcat gggccactgt ggcccgatgt ccccttcctt cccggccatt ctccctcttg
1141 ttttcacttc acttttcgtg ggatggtgtt accttagcta actaactctc ctccatgttg
1201 cctgtctttc ccagacttgt ccctcctttt ccagcgggac tcttctcatc cttcctcatt
1261 tgcaaggtga acacttcctt ctagggagca ccctcccacc ccccacccccc ccccacacac
1321 catctttcca tcccaggctt ttgaaaaaca aacagaaacc cgtgtatctg ggatatttcc
1381 atatggcaat aggtgtgaac agggaactca gaatacagac aagtagaaag attctcgctt
1441 aaaaaaatgt atttatttta tggcaagttg gaaaatatgt aactggaatc tcaaaagttc
1501 tttgggacaa aacagaagtc catggagtta tctaagctct tgtaagtgag ttaatttaaa
1561 aaagaaaatt aggctgagag cagtggctca cgcctgtaat cccagaactt tgggaggcta
1621 aggtgggtgg atcacctgag gtcaagagtt ccagaccagg ctggccagca tggtgaaacc
1681 ccgtctgtac taaaaataca aaaaattaac tgggcatggt agtgggtgcc tgtaatccca
1741 gctacttggg aggctgaggt gggagaattg ctcgaacctt ggaggtggag gttgtggtga
1801 gccatgatcg caccactgca ctctagcctg ggtgaccgag ggaggctctg tctcaaaagc
1861 aaagcaaaaa caaaaacaaa aacacctaaa aaacctgcag ttttgtttgt actttgtttt
1921 taaattatgc tttctatttt gagatcattg caaactcaac acaattgtaa gtaatgatac
1981 agagggatct tgtgtaccct tcacccagcc tcccccaatg gcaacatctt gcaaaactac
2041 aatgtagtct cataaccagg atattgacat tgatacagtg aagatacagg acattctcat
2101 caccacaggg atccccagga tgcccacttc cctccacccc cacaccccag ccgtgtccct
2161 aacccctggc aaccaggaat ccactctcca tttctataat gttgtcattt caagaatgtt
2221 attcaatgga atcatatagt atgtaacctg ttttgagctt aaaaaaaaaa gtatacatga
2281 ctttaatgag gaaaataaaa atgaatattg aaaaaaaaaa ctttagag
```

Figure 4

MNSFNYTTPDYGHYDDKDTLDLNTPVDKTSNTLRVPDILALVIFAVVFLVGVLGNAL
VVWVTAFEAKRTINAIWFLNLAVADFLSCLALPILFTSIVQHHHWPFGGAACSILPS
LILLNMYASILLLATISADRFLLVFKPIWCQNFRGAGLAWIACAVAWGLALLLTIPS
FLYRVVREEYFPPKVLCGVDYSHDKRRERAVAIVRLVLGFLWPLLTLTICYTFILLR
TWSRRATRSTKTLKVVVAVVASFFIFWLPYQVTGIMMSFLEPSSPTFLLLNKLDSLC
VSFAYINCCINPIIYVVAGQGFQGRLRKSLPSLLRNVLTEESVVRESKSFTRSTVDT
MAQKTQAV

Figure 5

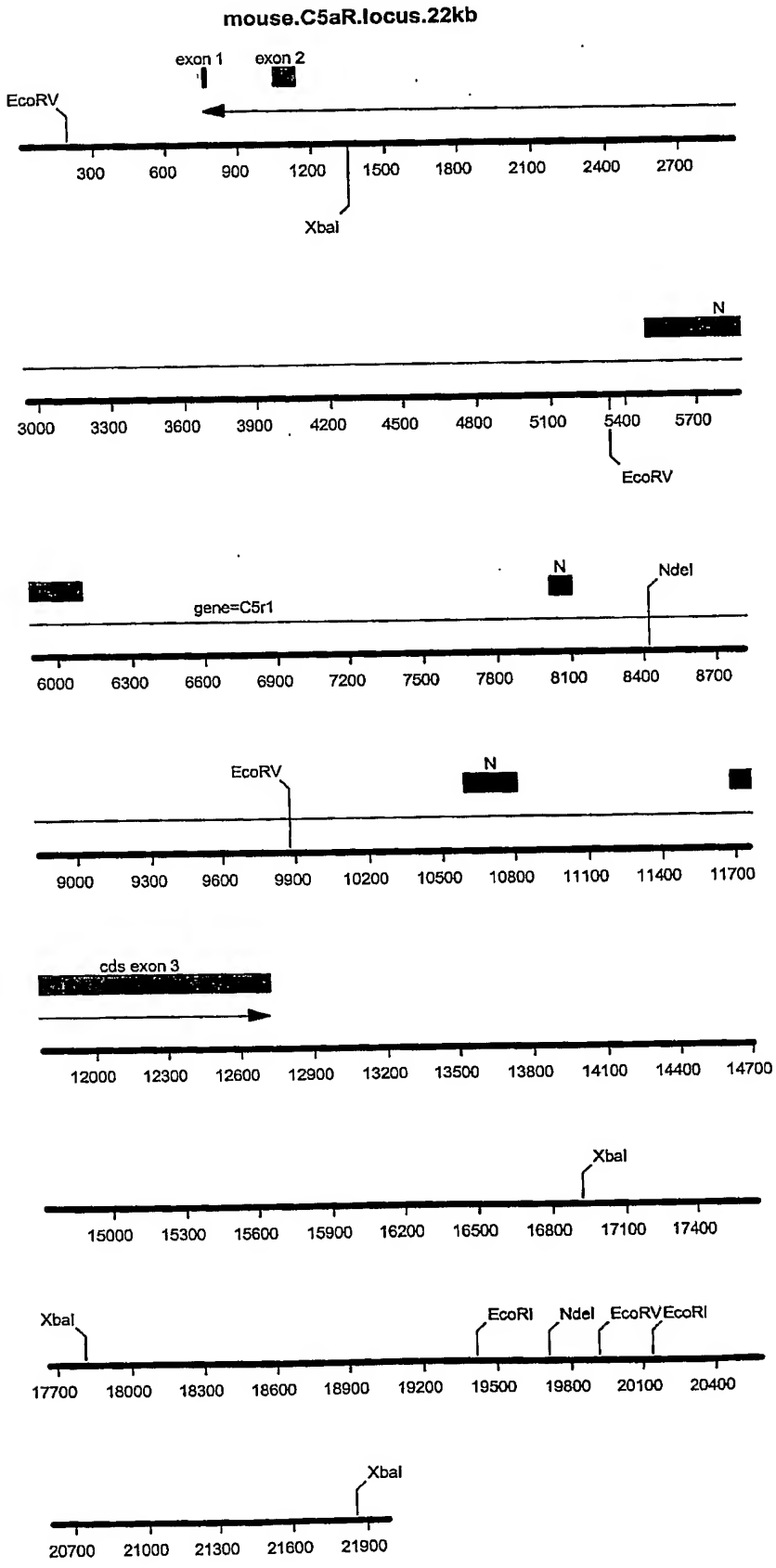
mouse.C5aR.locus.22kb
exon 1
exon 2
EcoRV
300
600
900
1200
1500
1800
2100
2400
2700
XbaI
N
3000
3300
3600
3900
4200
4500
4800
5100
5400
5700
EcoRV
N
NdeI
gene=C5r1
6000
6300
6600
6900
7200
7500
7800
8100
8400
8700
EcoRV
N
9000
9300
9600
9900
10200
10500
10800
11100
11400
11700
cds exon 3
12000
12300
12600
12900
13200
13500
13800
14100
14400
14700
XbaI
15000
15300
15600
15900
16200
16500
16800
17100
17400
XbaI
EcoRI
NdeI
EcoRV
EcoRI
17700
18000
18300
18600
18900
19200
19500
19800
20100
20400
XbaI
20700
21000
21300
21600
21900

Figure 6